The Bit Error Performance and Information Transfer Rate of SPAD Array Optical Receivers

The Bit Error Performance and Information Transfer Rate of SPAD Array Optical Receivers

Journal:	*IEEE Transactions on Communications*
Manuscript ID	TCOM-TPS-19-1459.R2
Manuscript Type:	Transactions Paper Submissions
Date Submitted by the Author:	16-Mar-2020
Complete List of Authors:	Sarbazi, Elham; The University of Edinburgh, Institute for Digital Communications Safari, Majid; University of Edinburgh, Institute for Digital Communication; Haas, Harald; The University of Edinburgh, Institute for Digital Communications School of Engineering and Electronics
Keyword:	Optical communication

The Bit Error Performance and Information Transfer Rate of SPAD Array Optical Receivers

Elham Sarbazi, *Student Member, IEEE*, Majid Safari, *Member, IEEE*, and Harald Haas, *Fellow Member, IEEE*

Abstract

In this paper the photon counting characteristics, the information rate and the bit error performance of single-photon avalanche diode (SPAD) arrays are investigated. It is shown that for sufficiently large arrays, the photocount distribution is well approximated by a Gaussian distribution with dead-time-dependent mean and variance. Because of dead time, the SPAD array channel is subject to counting losses, part of which are due to inter-slot interference (ISI) distortions. Consequently, this channel has memory. The information rate of this channel is assessed. Two auxiliary discrete memoryless channels (DMCs) are proposed which provide upper and lower bounds on the SPAD array information rate. It is shown that in sufficiently large arrays, ISI is negligible and the bounds are tight. Under such conditions, the SPAD array channel is precisely modelled as a memoryless channel. A discrete-time Gaussian channel with input-dependent mean and variance is adopted and the properties of the capacity-achieving input distributions are studied. Using a numerical algorithm, the information rate and the capacity-achieving input distributions, subject to peak and average power constraints are obtained. Furthermore, the bit error performance of a SPAD-based system with on-off keying (OOK) is evaluated for various array sizes, dead times and background count levels.

Index Terms

Single-photon avalanche diode (SPAD), SPAD array, photon counting, dead time, capacity, optical wireless communications.

I. Introduction

Single photon avalanche diodes (SPADs) are semiconductor devices with extremely high sensitivity, high power efficiency, high detection efficiency, and high timing resolution. They operate in avalanche breakdown mode, also known as the *Geiger mode*. In this mode, the SPAD is biased above its breakdown voltage such that individual photons trigger an avalanche breakdown, leading to a large internal gain and a measurable output current spike [1]. After each breakdown, a *quenching circuit* resets the SPAD by reducing the bias voltage below the breakdown threshold, stopping the avalanche and raising the bias voltage again. During the quenching process, the SPAD is unresponsive to incident photons for a circuit specific *dead time*. It is only after this dead time period that the SPAD is able to detect another photon. Active quenching (AQ) and passive quenching (PQ) are the two main types of quenching circuits. In AQ SPADs, the dead time is constant, while in PQ SPADs the dead time is extended by each incident photon. Thus, AQ SPADs generally have higher count rates than PQ SPADs [2], [3].

In recent years, SPADs have drawn particular attention in the field of optical wireless communication (OWC). Their unique features has opened the door to many OWC applications, dealing with very low intensity levels down to a single photon, where conventional positive-intrinsic-negative (PIN) diodes and avalanche photodiodes (APDs) can not provide sufficient sensitivity and power efficiency. Providing a very large internal gain and easily overcoming the thermal noise, SPADs are able to detect individual photons without the need for high-gain low-noise transimpedance amplifiers (TIAs). As a result of this, longer communication links can be supported with a SPAD receiver. However, the SPAD photon counting performance is degraded by its unavoidable dead time. In any SPAD-based OWC system, the counting losses arising from dead time not only result in higher error probabilities, but also limit the count rate, and hence the maximum achievable data rate. Efforts are still underway to improve the performance of these photodetectors by shortening the dead time period as much as possible. Meanwhile, several experimental studies suggest that the impact of dead time can be mitigated by employing arrays of SPADs [4]–[7].

Previous Works

Most of the related research articles are experimental studies [4]–[12] and the theoretical studies on the subject are limited. While the existing experimental results appreciate the presence and impact of dead time on the overall performance of the systems, there is a lack of an in-depth analysis and understanding of how exactly the dead time distorts the data reception and why

systems consisting of SPAD arrays are more robust to the dead time. In the analyses conducted in [13]–[25], the SPAD dead time has not been taken into consideration and an ideal Poisson process is assumed for the SPAD photon counting process. Authors in [26]–[29], have assumed a Poisson distribution with a dead-time-modified mean for the SPAD photon counting process. In [30], [31], a practical photon counting receiver in optical scattering communication (OSC) with finite sampling rate, paralyzable dead time, and electrical noise is considered. It is shown that the dead time leads to a sub-Poisson distribution for the number of recorded pulses. An approximate expression is derived for the photocount distribution which is only applicable for extremely low photon rates.

In [32] and [33], we presented a thorough characterization of single SPADs, where we derived the exact photocount distribution of both AQ and PQ single SPADs under the limits of a short dead time. We investigated the effect of a long dead time on the photocount statistics of AQ single SPADs in [34]. We also studied the information transfer rate of AQ single SPADs with binary signalling in [35] where the SPAD was modelled as a discrete memoryless channel (DMC). In [36], we provided an initial statistical modelling and error performance evaluation of SPAD arrays with relatively short dead times.

Our Contribution

In this paper, we focus on SPAD arrays, as they can tolerate longer dead times and offer higher count rates, hence, higher data rates. We extend our earlier work of [36] and characterize an AQ SPAD array for OWC applications. We derive the photocount distribution of SPAD arrays and employ Monte Carlo methods to verify the validity of the analytical models. In particular, we consider inter-slot interference (ISI) distortions introduced by the SPAD dead time. We assess the impact of dead time on the information transfer rate of SPAD array. We also evaluate the bit error ratio (BER) of an OWC system under the limits of dead time and examine the ISI effects.

The remainder of this paper is organized as follows. In Section II, the photocount statistics of single SPADs and SPAD arrays are discussed. In Section III, the communication channel model of the SPAD array is presented. In Section IV, the information rate of SPAD arrays is investigated. In Section V the bit error performance of a SPAD-based OWC system is evaluated. Finally, concluding remarks are provided in Section VI.

II. Photocount Statistics

A. Single SPAD

Assume that the photons arrive on the surface of the SPAD detector according to a Poisson arrival process. After each photon detection, the SPAD becomes unresponsive for the duration of its dead time and can not detect any other photons. This leads to some *counting losses*. In the photon counting context, the *photocount* refers to the number of successfully detected photons in the so-called *counting interval*. The number of counting losses is random and depends on the *relative* dead time, i.e., the dead time to the counting interval ratio.

If the SPAD is free at the beginning of the counting interval, the first incident photon is indeed detected. However, in many cases of practical interest, the SPAD may be blocked for some time after the counting interval has started. In such cases, the first few incident photons may be missed as well, leading to higher counting losses. Depending on the presence or absence of dead time, the photocount statistics of a single SPAD can be summarized as follows:

A.1. Without Dead Time

Consider a counting interval of length T and a constant average photon rate λ (in photons/s). Let the random variable K denote the photocounts in the counting interval and k be an arbitrary realization of K. For an ideal single SPAD without dead time, the photocounts follow a Poisson distribution [37]:

$$p_0(k) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}. \tag{1}$$

The mean and variance of the photocounts are given by [37]:

$$\mu_0 = \lambda T, \tag{2a}$$

$$\sigma_0^2 = \lambda T. \tag{2b}$$

A.2. With Dead Time: Free at the Beginning of the Counting Interval

Assume that the SPAD is ready to operate at the beginning of the counting interval, such that the first incident photon is certainly detected. With a dead time of duration τ, the maximum number of photocounts is [32], [33], [38]:

$$k_{\max} = \lfloor T/\tau \rfloor + 1, \tag{3}$$

where $\lfloor x \rfloor$ denotes the largest integer smaller than x. In this case, the photocounts are no longer Poisson distributed and the probability mass function (PMF) of the photocounts is given by [32], [33], [38]:

$$p_K(k) = \begin{cases} \sum_{i=0}^{k} \psi(i, \lambda_k) - \sum_{i=0}^{k-1} \psi(i, \lambda_{k-1}), & k < k_{\max} \\ 1 - \sum_{i=0}^{k-1} \psi(i, \lambda_{k-1}), & k = k_{\max} \\ 0, & k > k_{\max} \end{cases} \tag{4}$$

where the function $\psi(i, u)$ is defined as $\psi(i, u) = u^i e^{-u}/i!$ and $\lambda_k = \lambda(T - k\tau)$. The mean and variance of the above PMF are expressed as [32], [33]:

$$\mu_K = k_{\max} - \sum_{k=0}^{k_{\max}-1} \sum_{i=0}^{k} \psi(i, \lambda_k), \tag{5a}$$

$$\sigma_K^2 = \sum_{k=0}^{k_{\max}-1} \sum_{i=0}^{k} (2k_{\max} - 2k - 1)\psi(i, \lambda_k) - \left(\sum_{k=0}^{k_{\max}-1} \sum_{i=0}^{k} \psi(i, \lambda_k) \right)^2. \tag{5b}$$

If $\tau \ll T$, the expressions in (5a) and (5b) can be approximated as [33]:

$$\mu_K \approx \frac{\lambda T}{1 + \lambda\tau}, \tag{6a}$$

$$\sigma_K^2 \approx \frac{\lambda T}{(1 + \lambda\tau)^3}. \tag{6b}$$

A.3. With Dead Time: Blocked at the Beginning of the Counting Interval

Assume that the SPAD is blocked for ρ seconds at the beginning of the counting interval. In this case, the probability of detecting k photons is given by:

$$p_K(k) = \begin{cases} \sum_{i=0}^{k} \psi(i, \lambda'_k) - \sum_{i=0}^{k-1} \psi(i, \lambda'_{k-1}), & k < k'_{\max} \\ 1 - \sum_{i=0}^{k-1} \psi(i, \lambda'_{k-1}), & k = k'_{\max} \\ 0, & k > k'_{\max} \end{cases} \tag{7}$$

where $k'_{\max} = \lfloor (T - \rho)/\tau \rfloor + 1$ and $\lambda'_k = \lambda(T - \rho - k\tau)$. The mean and variance are expressed as:

$$\mu_K = k'_{\max} - \sum_{k=0}^{k'_{\max}-1} \sum_{i=0}^{k} \psi(i, \lambda'_k), \tag{8a}$$

$$\sigma_K^2 = \sum_{k=0}^{k'_{\max}-1} \sum_{i=0}^{k} (2k'_{\max} - 2k - 1)\psi(i, \lambda'_k) - \left(\sum_{k=0}^{k'_{\max}-1} \sum_{i=0}^{k} \psi(i, \lambda'_k) \right)^2. \tag{8b}$$

B. SPAD Array

SPAD arrays are more robust to the dead time counting losses and can achieve higher count rates. The output of a SPAD array is the sum of all photocounts from individual SPADs during the same time interval. Other than the dead time, the fill factor (FF) of the array also affects the photocount distribution. The FF is defined as the ratio of the SPAD active area to the total array area and is denoted by a coefficient $C_{\rm FF}$. Throughout this paper, without loss of generality, we assume that $C_{\rm FF} = 1$, i.e., the entire surface of the SPAD is sensitive and therefore, the probability that incident photons hit the active area is equal to one.

Consider a SPAD array consisting of $N_{\rm array}$ elements, and assume independent counting statistics for the single SPADs in the array (due to negligible crosstalk[1]). Denoting by K_i the photocount at the ith SPAD, the array output is expressed as:

$$Y = \sum_{i=1}^{N_{\rm array}} K_i. \tag{9}$$

If $N_{\rm array}$ is sufficiently large, according to the central limit theorem (CLT) the photocount distribution of the SPAD array can be approximated by a Gaussian distribution[2], that is:

$$p_Y(y) \sim \mathcal{N}(\mu_Y, \sigma_Y^2), \tag{10}$$

where,

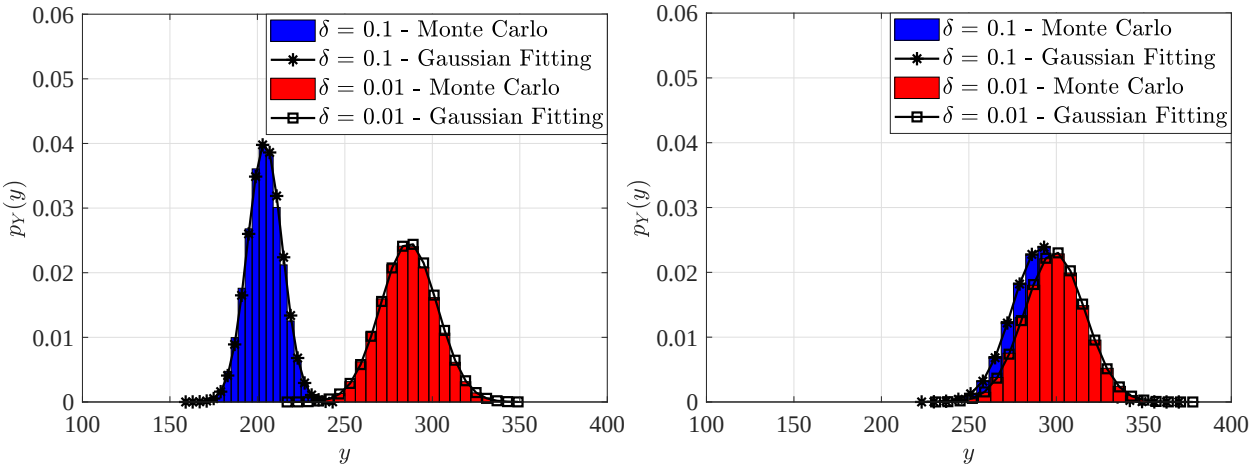
$$\mu_Y = \sum_{i=1}^{N_{\rm array}} \mu_{K_i}, \quad \sigma_Y^2 = \sum_{i=1}^{N_{\rm array}} \sigma_{K_i}^2. \tag{11}$$

Here, μ_{K_i} and $\sigma_{K_i}^2$ are the mean and variance of the photocount distribution of the ith SPAD in the array, respectively.

The approximate counting distribution given in (10) is compared with the Monte Carlo simulation results in Fig. 1 for a free SPAD array. It is assumed that all the array elements are identical. As shown, the Monte Carlo simulation results and the Gaussian approximations

[1]Crosstalk is a phenomenon that takes place in SPAD arrays, when the avalanche in one SPAD triggers an undesired secondary avalanche in a neighbouring SPAD.

[2]Our extensive simulations show that for $N_{\rm array} > 10$ this statement is valid.



(a) $N_{\text{array}} = 64$ (b) $N_{\text{array}} = 1024$

Fig. 1. Probability distribution of photocounts for a free SPAD array with $\lambda T = 300$ photons, $C_{\text{FF}} = 1$.

are perfectly matched and this confirms the validity of the approximation. In Fig. 1a, an array of 64 SPADs and in Fig. 1b, a larger array with 1024 SPADs are considered. Let the dead time ratio (i.e., the normalized dead time) be defined as:

$$\delta = \frac{\tau}{T}. \tag{12}$$

It is observed that for $\delta = 0.01$ the arrays have similar histograms, meaning that the effect of dead time is eliminated in both arrays. However, for $\delta = 0.1$ the array with 1024 SPADs has a higher mean value, i.e., the counting losses arising from the longer dead time are effectively mitigated. Therefore, larger arrays are more robust to dead time losses.

III. Communication Channel Model of the SPAD Array

Some applications may require the SPAD receiver to operate in consecutive counting intervals. Of particular interest are SPAD-based OWC systems. In these systems, the SPAD dead time has two effects on the photon counting process. The primary effect is to cause some counting losses in each time slot (e.g., symbol time). The secondary effect is incurring counting losses in the neighbouring intervals, where the dead time of last detected photon in one interval may overlap with the next interval(s), leading to a temporal blockage of the SPAD at the beginning of the time intervals, thereby causing extra counting losses and distorting the output (the photocounts). We refer to this as ISI distortions.

Consider the SPAD array as a communication channel and assume a discrete-time signalling scheme: the channel input is the intensity of the optical signal which can vary between discrete

time slots of length T while remaining constant within each time interval. The channel output is the SPAD array photocount in each time interval, corrupted by background counts of a constant intensity $\lambda_{\rm b}$ photons/s. Thus, $K_{\rm b} = \lambda_{\rm b} T$ is the average background counts per counting interval. Here, we refer to all the noisy counts (arising from dark counts, afterpulsing, and ambient light) as the background counts. Let the random variables X_n and Y_n denote the channel input and output in the nth time interval, respectively. Further, x_n and y_n are arbitrary realizations of X_n and Y_n, respectively. We also use the notation $r^i \triangleq [r_1, r_2, \ldots, r_i]$ to refer to sequences where r can be replaced by X, x, Y and y.

In the presence of dead time and due to the ISI effect, the SPAD array channel has *memory*. Unlike classical memory channels in which y_n depends on previous inputs (x^{n-1}) and/or previous outputs (y^{n-1}), in the SPAD channel y_n depends on the *time of last photon detection events of individual SPADs in the previous interval*. Define:

$$\boldsymbol{\rho}_n = [\rho_n^1, \rho_n^2, \ldots, \rho_n^{N_{\rm array}}], \tag{13}$$

with ρ_n^m denoting the ISI blockage of the mth array element in the nth time interval given by:

$$\rho_n^m = \begin{cases} t_{n-1}^m - (n-1)T + \tau, & t_{n-1}^m > (n-1)T - \tau \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

where t_{n-1}^m is the time of last photon count of the mth array element in the $(n-1)$th time interval. Therefore, according to Section II-B, the following conditional probability describes the communication channel model of a SPAD array:

$$\Pr\{y_n | x_n, \boldsymbol{\rho}_n\} = \frac{1}{\sqrt{2\pi\sigma^2_{Y_n|X_n,\boldsymbol{\rho_n}}}} \exp\left[\frac{-(y_n - \mu_{Y_n|X_n,\boldsymbol{\rho_n}})^2}{2\sigma^2_{Y_n|X_n,\boldsymbol{\rho_n}}}\right], \quad x_n \in \mathbb{R}^+ \text{ and } y_n \in \mathbb{Z}^+ \tag{15}$$

In (8a) and (8b), the mathematical expressions of μ_K and σ_K^2 are functions of λ and ρ. We can define two multivariable functions $A(\lambda, \rho) = \mu_K$ and $B(\lambda, \rho) = \sigma_K^2$. Then:

$$\mu_{Y_n|X_n,\boldsymbol{\rho_n}} = \sum_{m=1}^{N_{\rm array}} A\left(\frac{x_n + \lambda_{\rm b}}{N_{\rm array}}, \rho_n^m\right), \tag{16a}$$

$$\sigma^2_{Y_n|X_n,\boldsymbol{\rho_n}} = \sum_{m=1}^{N_{\rm array}} B\left(\frac{x_n + \lambda_{\rm b}}{N_{\rm array}}, \rho_n^m\right). \tag{16b}$$

Based on (15), (16a) and (16b), the SPAD array channel is clearly a *discrete-time Gaussian channel* with *signal-dependent mean and variance*. This channel, however, is not memoryless.

A finite-state channel (FSC) model[3] can be used to characterize this channel. Let S_n denote the state at time n with the state space $\mathcal{S} = \{1, 2, \ldots, N_S\}$. The state space $\mathcal{S}$ corresponds to N_S discrete memoryless channels, with common input and output alphabet sets. Assume that the channel inputs are independent of its states. Define the following conditional state probabilities:

$$\alpha_n^i = \Pr\{S_n = i | y^{n-1}\}, \tag{17}$$

$$\beta_n^i = \Pr\{S_n = i | x^{n-1}, y^{n-1}\}. \tag{18}$$

Some of the main properties of this FSC are listed in Appendix A. For an FSC with these properties, the following theorem holds:

Theorem 1. *For independent and identically distributed (iid) inputs, the N_S-dimensional random vectors $\boldsymbol{\alpha}_n = [\alpha_n^1, \alpha_n^2, \ldots, \alpha_n^{N_S}]$ and $\boldsymbol{\beta}_n = [\beta_n^1, \beta_n^2, \ldots, \beta_n^{N_S}]$ are Markov chains that converge to steady distributions which are independent of the initial channel state [40].*

This theorem is a direct result of the FSC being stationary and ergodic. This FSC has a unique stationary distribution regardless of the initial state distributions. The stationary distribution is also known as the *equilibrium* or *steady state* distribution.

Although the FSC model for the SPAD array channel accurately represents the properties of the channel memory, it is mathematically intractable when it comes to information rate analysis. Therefore, in the following, we introduce two auxiliary DMCs. We will use these channels in Section IV for the information rate analysis.

A. Auxiliary DMCs

The first auxiliary channel is a DMC in which the ISI is ignored. More precisely, this DMC is only subject to the primary dead time losses. So the PMF, mean and variance given in (4), (5a) and (5b), respectively, describe the photocount statistics of each SPAD in the array. Recall from Section II-A2, that the single SPADs are assumed to be free at the beginning of each counting interval (i.e., time slots of the communication system), thus these expressions do not take the effect of ISI impairments into account. Hereinafter, this channel is referred to as DMC_1. The

[3]An FSC is a discrete-time channel for which the distribution of the channel output depends on both the input and the underlying channel state [39]. This allows the output to have an implicit dependence on previous inputs and outputs via the channel state.

following conditional probability describes the communication channel model of DMC_1:

$$\Pr\{y_n|x_n\} = \frac{1}{\sqrt{2\pi\sigma^2_{Y_n|X_n}}} \exp\left[\frac{-(y_n - \mu_{Y_n|X_n})^2}{2\sigma^2_{Y_n|X_n}}\right], \quad x_n \in \mathbb{R}^+ \text{ and } y_n \in \mathbb{Z}^+ \tag{19}$$

with

$$\mu_{Y_n|X_n} = \sum_{m=1}^{N_{\text{array}}} A_1\left(\frac{x_n + \lambda_{\text{b}}}{N_{\text{array}}}\right), \tag{20a}$$

$$\sigma^2_{Y_n|X_n} = \sum_{m=1}^{N_{\text{array}}} B_1\left(\frac{x_n + \lambda_{\text{b}}}{N_{\text{array}}}\right), \tag{20b}$$

where $A_1(\lambda) = \mu_K$ as given in (5a) and $B_1(\lambda) = \sigma^2_K$ as given in (5b).

The second auxiliary channel is a DMC, denoted by DMC_2, in which there is an output power degradation that arises due to the memory introduced by the ISI distortion. This is inspired by the work presented in [41]–[43]. Recall from Theorem 1 that the SPAD Markov process reaches equilibrium. Further, the photocount statistics of consecutive intervals asymptotically become independent in the steady state. Therefore, we can adopt the steady state distribution for the DMC_2. This reflects the effects of both primary dead time losses and the average ISI. In this paper, the steady state distributions are obtained through Monte Carlo simulations, as the analytical expressions are unwieldy. Comprehensive details of the Monte Carlo simulations can be found in [33].

In the next section, we will show that the auxiliary channel DMC_1 provides an upper bound and DMC_2 gives a lower bound on the information rate of the SPAD array.

IV. Information Transfer Rate Analysis

In this section, the information transfer rates of the SPAD array channel and the auxiliary DMCs are investigated.

A. Information Rate Analysis for the SPAD Array Channel

As discussed in Section III, the SPAD Markov chain is stationary and ergodic over the finite state space $\mathcal{S}$. Therefore, the effect of its initial state dies away as n grows and the FSC is indecomposable [39]. The capacity of an indecomposable FSC is independent of its initial state, and is given by [39]:

$$C = \lim_{n\to\infty} \max_{P_{X^n}(x^n)} \frac{1}{n} I(X^n; Y^n) \tag{21}$$

where $P_{X^n}(x^n)$ denotes the input distribution on X^n. Due to practical considerations and device limitations, such as the saturation of SPAD receivers at high intensities, the input signal is often subject to peak and average power constraints. Also, since X_n is the light intensity, the constraints are directly imposed on X_n. In addition, X_n should be nonnegative. Thus, the constraints are:

$$0 \leq X_n \leq \mathcal{A}, \tag{22a}$$

$$\mathbb{E}[X_n] \leq \mathcal{E}, \tag{22b}$$

where $\mathcal{A}$ and $\mathcal{E}$ are the peak and average powers, respectively. Without loss of generality, it is assumed that $0 \leq \mathcal{E} \leq \mathcal{A}$ and $\mathcal{A}$ is finite.

The mathematical simplification of the capacity expression given in (21) with general input distributions is intractable. In this study, we focus on iid inputs. Therefore, the maximization is performed over all admissible input distributions satisfying the constraints given in (22), for which $P_{X^n}(x^n) = \prod_{i=1}^{n} P_X(x_i)$. Denote the resulting "maximum information rate" by I_{iid}. It is evident that $I_{\text{iid}} < C$. Despite the assumption of iid X_i's, the mathematical simplification and the numerical calculation of I_{iid} still remain cumbersome and as a result, we resort to establishing lower and upper bounds on I_{iid}. The following theorem states that the maximum achievable information rate of the SPAD array channel with iid inputs is bounded by the capacities of the two DMC channels introduced in Section III-A.

Theorem 2. *The maximum information rate I_{iid} is bounded as:*

$$C_{\text{DMC}_2} \leq I_{\text{iid}} \leq C_{\text{DMC}_1} \tag{23}$$

where C_{DMC_1} and C_{DMC_2} are the capacities of DMC$_1$ and DMC$_2$, respectively, and the equalities hold if and only if $\delta = 0$. Hereinafter, we use $I_{\text{U}} = C_{\text{DMC}_1}$ and $I_{\text{L}} = C_{\text{DMC}_2}$ to refer to these bounds.

Proof: Please refer to Appendix B. ■

B. Capacity Analysis for the Auxiliary DMCs

As discussed in Section III-A, the mean and variances of the DMCs are *signal-dependent*, unlike the classical Gaussian channels [44]. Such a class of Gaussian channels whose conditional output distribution given the channel input is Gaussian with input-dependent mean and variance are termed as conditionally Gaussian (CG) channels [45], [46]. Although the properties of such channels have been studied in the literature, their capacity is not yet known [45], [46].

Nevertheless, it is well known that subject to peak and average power constraints, the channel capacity is achievable and the capacity-achieving distribution is unique and discrete with a finite number of mass points for finite A and $\mathcal{E}$ [45], [46]. In what follows, some of the findings in the aforementioned reference articles are summarized and adopted for studying the capacity of the auxiliary DMCs.

Assume an input distribution defined over constellation $\psi_x = \{a_1, a_2, \ldots, a_l\}$, with probability distribution $\psi_p = \{p_1, p_2, \ldots, p_l\}$, where $l = \|\psi_x\|$ and $0 \leq a_1 < a_2 < \cdots < a_l \leq \mathcal{A}$. Denote by P_X the corresponding cumulative distribution function (CDF), that is:

$$dP_X = p_1\delta(x - a_1) + p_2\delta(x - a_2) + \cdots + p_l\delta(x - a_l). \tag{24}$$

where $\delta(\cdot)$ is the Dirac delta function. Also, let $\mathcal{P}_X$ be the set of all input distributions satisfying the constraints defined in (22):

$$\mathcal{P}_X \triangleq \left\{ P_X : \int_0^{\mathcal{A}} dP_X = 1, \mathbb{E}[X] \leq \mathcal{E} \right\}. \tag{25}$$

For the DMC channel under consideration, let $p_{Y|X}(y|x)$ be the conditional probability of Y given X. For each P_X, denote the corresponding distribution of Y by $p_Y(y; P_X)$, the marginal entropy of Y by $H(Y; P_X)$, the conditional entropy of Y given X by $H(Y|X; P_X)$, and the mutual information between Y and X by $I(P_X)$ [47]:

$$p_Y(y; P_X) = \int_x p_{Y|X}(y|x) dP_X \tag{26}$$

$$H(Y; P_X) = -\sum_y p_Y(y; P_X) \log_2 p_Y(y; P_X) \tag{27}$$

$$H(Y|X; P_X) = \frac{1}{2} \int_x \log_2 2\pi e\sigma^2 \, dP_X \tag{28}$$

$$I(P_X) = \int_x \left[\sum_y p_{Y|X}(y|x) \log_2 \frac{p_{Y|X}(y|x)}{p_Y(y; P_X)} \right] dP_X \tag{29}$$

And the channel capacity is [47]:

$$C = \max_{P_X \in \mathcal{P}_X} I(P_X) \, . \tag{30}$$

Let the capacity-achieving values of ψ_x, ψ_p, and P_X subject to constraints $\mathcal{A}$ and $\mathcal{E}$, be denoted by $\psi_x^*(\mathcal{A}, \mathcal{E})$, $\psi_p^*(\mathcal{A}, \mathcal{E})$, and $P_X^*(\mathcal{A}, \mathcal{E})$, respectively. Some of the main properties of the capacity-achieving distribution are as follows:

- *Existence and uniqueness*: There exists a unique probability measure P_X^* satisfying the bounded-input and average power constraints which maximizes $I(P_X)$ [45, Theorem 1].

Algorithm 1 Search algorithm for finding the capacity-achieving input distribution.

Input: $\mathcal{A}$, $\mathcal{E}$
Output: C, P_X^*

procedure CAPACITY($\mathcal{A}, \mathcal{E}$)
  $l \leftarrow 2$
  Solve (30) under the constraints that $\|\psi_x\| = l$ and $a_1 = 0$. ▷ See [44].
  Determine $\epsilon^{(l)}$ according to (33).
  if $\epsilon^{(l)} < 0$ **then**
    $l \leftarrow l + 1$
    go to 3
  end if
  if (31) holds for all $x \in [0, \mathcal{A}]$ **then**
    return C and P_X^*
  else
    $l \leftarrow l + 1$
    go to 3
  end if
end procedure

- *Necessary and sufficient condition*: P_X^* is capacity-achieving if and only if there exists $\epsilon \geq 0$ such that for all $x \in [0, \mathcal{A}]$ [45, Theorem 2]:

$$Q(x; P_X^*) - I(P_X^*) - \frac{1}{2}\log_2 2\pi e\sigma^2 - \epsilon(x - \mathcal{E}) \leq 0, \tag{31}$$

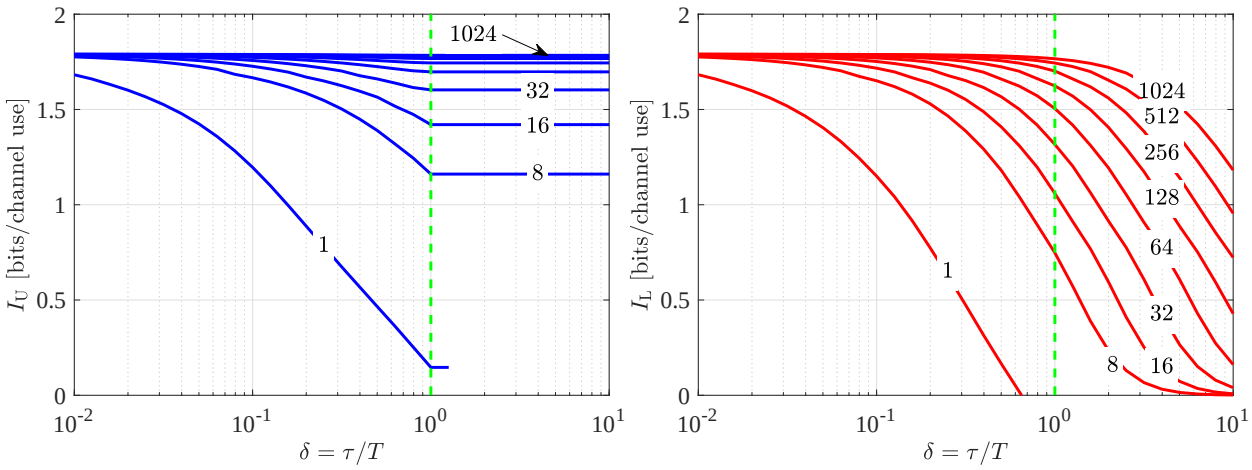
  where

$$Q(x; P_X) \triangleq -\sum_y p_{Y|X}(y|x) \log_2 p_Y(y; P_X). \tag{32}$$

- *Discreteness*: The capacity-achieving distribution P_X^*, is discrete and consists of a finite set of mass points [45, Theorem 3].
- *Mass point at zero*: The capacity-achieving distribution always contains a mass point located at zero [45, Proposition 1]. That is, $0 \in \psi_x^*(\mathcal{A}, \mathcal{E})$. Therefore,

$$\epsilon = \frac{1}{\mathcal{E}}\left[I(P_X^*) - Q(0; P_X^*) + \frac{1}{2}\log_2[2\pi e\sigma^2]\right]. \tag{33}$$

Although the above properties of the capacity-achieving distributions for the CG channels are known, closed-form analytical expressions for the capacity are unknown in general. Therefore, numerical methods are applied in order to compute the capacity and capacity-achieving distributions of the CG channels. We also follow a similar approach and obtain the optimal input distribution and the capacity of the auxiliary DMCs via the algorithm presented in Algorithm 1. This algorithm is adopted from [45].



(a) Upper bounds (b) Lower bounds

Fig. 2. Bounds on the information rate of the SPAD array for several array sizes: $\mathcal{A} = 50$, $\mathcal{E} = 20$ and $K_{\mathrm{b}} = 5$.

In Algorithm 1, the inputs are $\mathcal{A}$ and $\mathcal{E}$. The algorithm initializes with a binary distribution $(l = 2)$. In each iteration, first the optimal P_X which maximizes $I(P_X)$ is obtained using the method presented in [44]. More details can be found in [48], [49]. Since a mass point at $x = 0$ always exits, $\epsilon^{(l)}$ is determined using (33). Failure of the necessary condition $\epsilon^{(l)} > 0$ indicates that this P_X is not optimal and the current number of mass points, l, is not sufficient. Thus, the number of mass points should be increased by one, and the distribution function P_X which maximizes the information rate (subject to constraints) should be determined again. If $\epsilon^{(l)} > 0$, then the necessary and sufficient condition in (31) is tested. If it is satisfied, then P_X is the capacity-achieving probability measure. Otherwise, l is increased by one and the procedure is repeated.

C. Numerical Results and Discussions

In the following, some numerical results are provided for the above analysis. Note that the mean and variance of the SPAD array photocounts and both of the DMC channels are signal-dependent due to the dead time. Therefore, the dead time is the parameter that determines the degree of signal dependency.

Fig. 2 illustrates the bounds on the SPAD array maximum information rate as a function

of the dead time ratio δ for $A = 50$, $\mathcal{E} = 20$, $K_{\rm b} = 5$, and various array sizes[4]. As seen in Fig. 2a, the upper bounds remain constant for $\delta \geq 1$. According to (4), for $\delta \geq 1$, each SPAD in the array can detect at most one photon ($k_{\max} = 1$) and the photocounts follow a Bernoulli distribution. Accordingly, other than limiting $k_{\max}$, the dead time does not affect the array photocount distribution, and therefore the upper bounds remain constant for δ values larger than 1. As shown in Fig. 2, for $\delta \leq 0.1$, the bounds are tight for all the array sizes. This means that in all the arrays, the ISI distortion is negligible in this range. In addition, the bounds remain tight for $0.1 < \delta \leq 1$ if $N_{\rm array} \geq 256$.

Fig. 3 shows the bounds for $N_{\rm array} = 64$, as a function of the peak power constraint, $\mathcal{A}$, in Figs. 3a and 3b, and as a function of the average power constraint, $\mathcal{E}$, in Figs. 3c and 3d. In these figures, two different background noise levels ($K_{\rm b} = 5$ and 10) and two dead time ratios ($\delta = 0.1$ and 1) are considered. From Figs. 3a and 3c, it is seen that for $\delta = 0.1$, the bounds are tight, confirming the negligible effect of ISI. However, according to Figs. 3b and 3d, with $\delta = 1$, ISI is significant, leading to a gap between the bounds.

In Fig. 4, the bounds are provided for the array of $N_{\rm array} = 1024$ considering two different background noise levels and dead time ratios. As observed in this figure, the bounds are tight for $N_{\rm array} = 1024$ for all the parameter values. According to our extensive numerical investigation of the bounds, the following conclusions can be drawn.

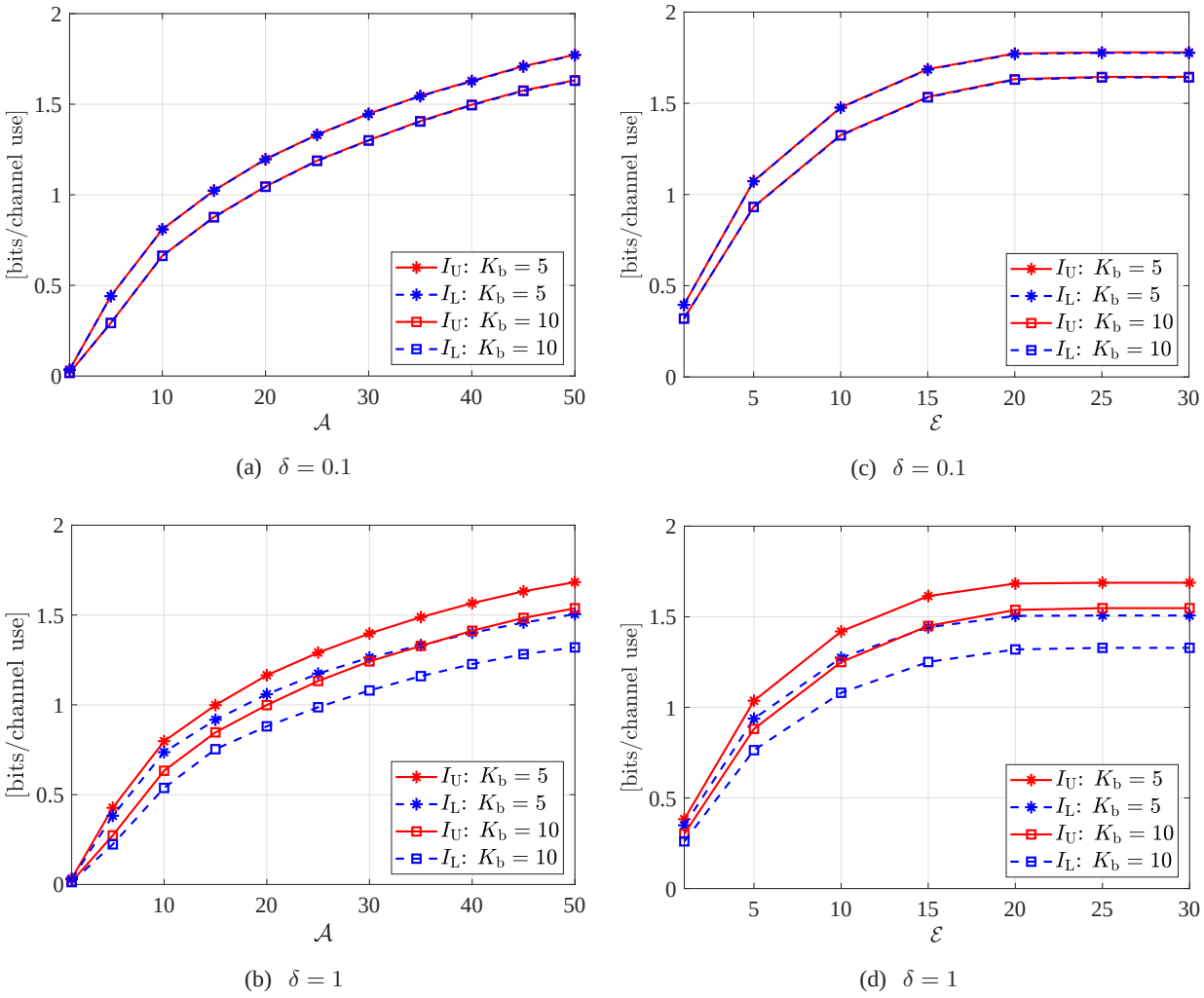
The maximum achievable information rates of:

- the SPAD array of $N_{\rm array} = 64$ with $\delta \leq 0.1$,
- the SPAD array of $N_{\rm array} = 1024$ with $\delta \leq 1$,

can be accurately approximated by their bounds. This means that in these cases, the SPAD array can be well approximated by the auxiliary DMCs. Figs. 2–4 support these statements. In the following numerical results, this approximation is used and the effect of different parameters on the capacity and the capacity-achieving input distribution are investigated.

In Fig. 5 the effect of peak power constraint, $\mathcal{A}$, and the average power constraint, $\mathcal{E}$, on the optimal mass points (ψ_x^*) and the corresponding probability measure (ψ_p^*) is presented for the array of 64 SPADs. In these figures, the parameters δ and $K_{\rm b}$ are assumed to remain fixed as follows: $\delta = 0.1$, $K_{\rm b} = 5$. Similarly, in Fig. 6, ψ_x^* and ψ_p^* are provided for an array of 1024

[4]As a practical example of these values, at typical 500 lux ambient light conditions, the background intensity level reaching a 1 mm^2 detector is around $\lambda_{\rm b} = 1.5 \times 10^9$ counts per second. This corresponds to $K_{\rm b} = 15$ photons per symbol, assuming on-off keying (OOK) modulation with a bit interval of $T = 10$ ns. Also, a signal average count of $K_{\rm s} = 50$ corresponds to a received optical power of 3 mW/m^2 at a wavelength of 400nm.



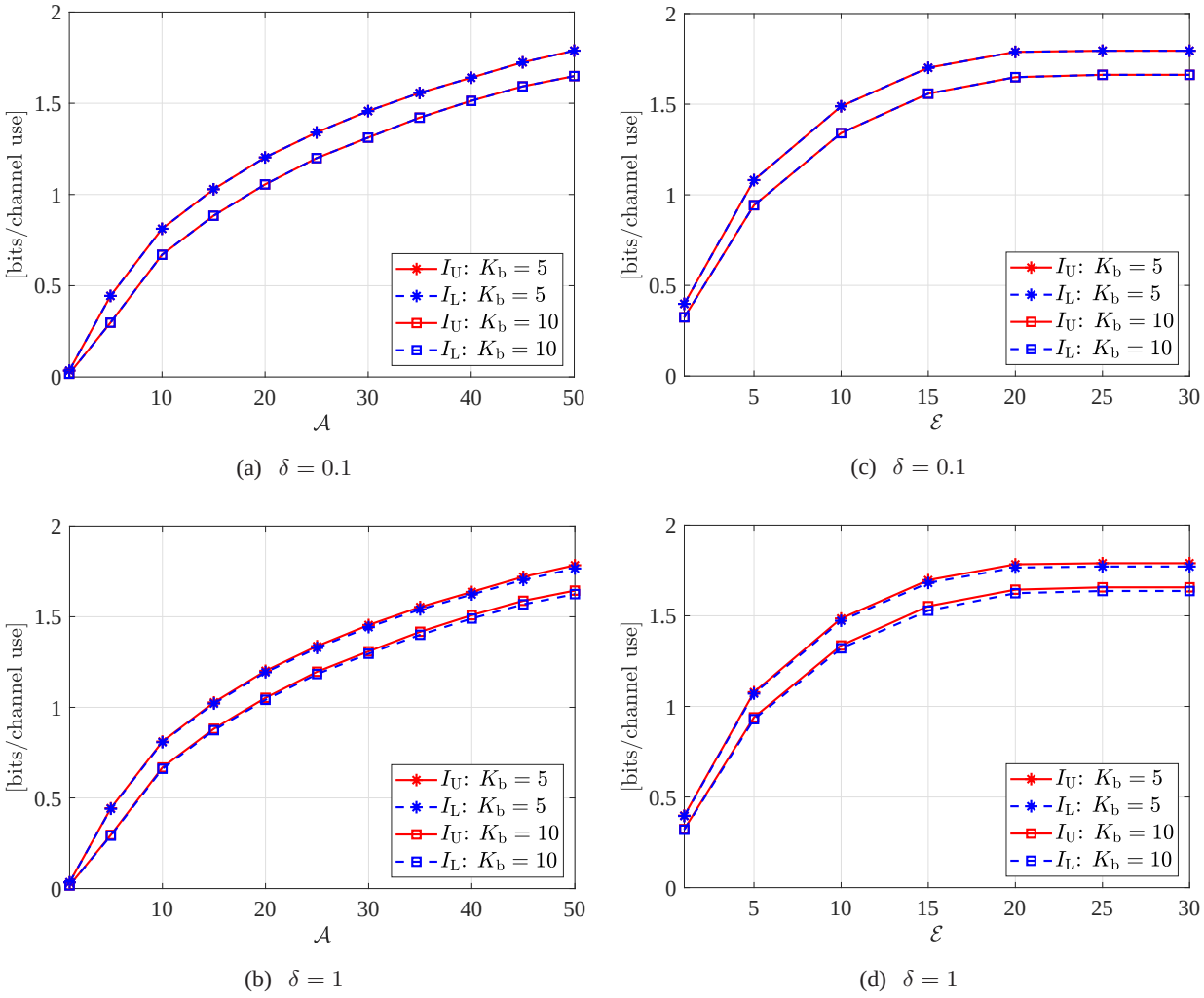
(a) $\delta = 0.1$ (c) $\delta = 0.1$

(b) $\delta = 1$ (d) $\delta = 1$

Fig. 3. Bounds on the information rate of SPAD array for $N_{\mathrm{array}} = 64$: (a),(b) $\mathcal{E} = 20$; (c),(d) $\mathcal{A} = 50$.

SPADs with $\delta = 1$ and $K_{\mathrm{b}} = 10$. In Fig. 7 the effect of dead time on ψ_x^* and ψ_p^* is shown for the array of $N_{\mathrm{array}} = 1024$. From Figs. 5–7, the following remarks are deduced:

- The capacity-achieving measure (ψ_x^*, ψ_p^*) contains two mass points, one at $x = 0$ and one at $x = \mathcal{A}$, for all the parameter values. In [45], it is proved that $x = 0$ is always a mass point of (ψ_x^*, ψ_p^*). However, it is not proved whether $x = \mathcal{A}$ is also a mass point in general.
- As $\mathcal{A}$ increases, more mass points are required for achieving the capacity. In the presented range of parameters, the variance of the signal-dependent noise term is almost a linear function of the signal power. As a result, for higher values of $\mathcal{A}$, the average signal-dependent noise power is higher; hence, a signalling scheme with a larger constellation size is more favourable.
- As $\mathcal{E}$ increases, the capacity also increases almost until $\mathcal{E} \approx 0.5\mathcal{A}$. After this point, the



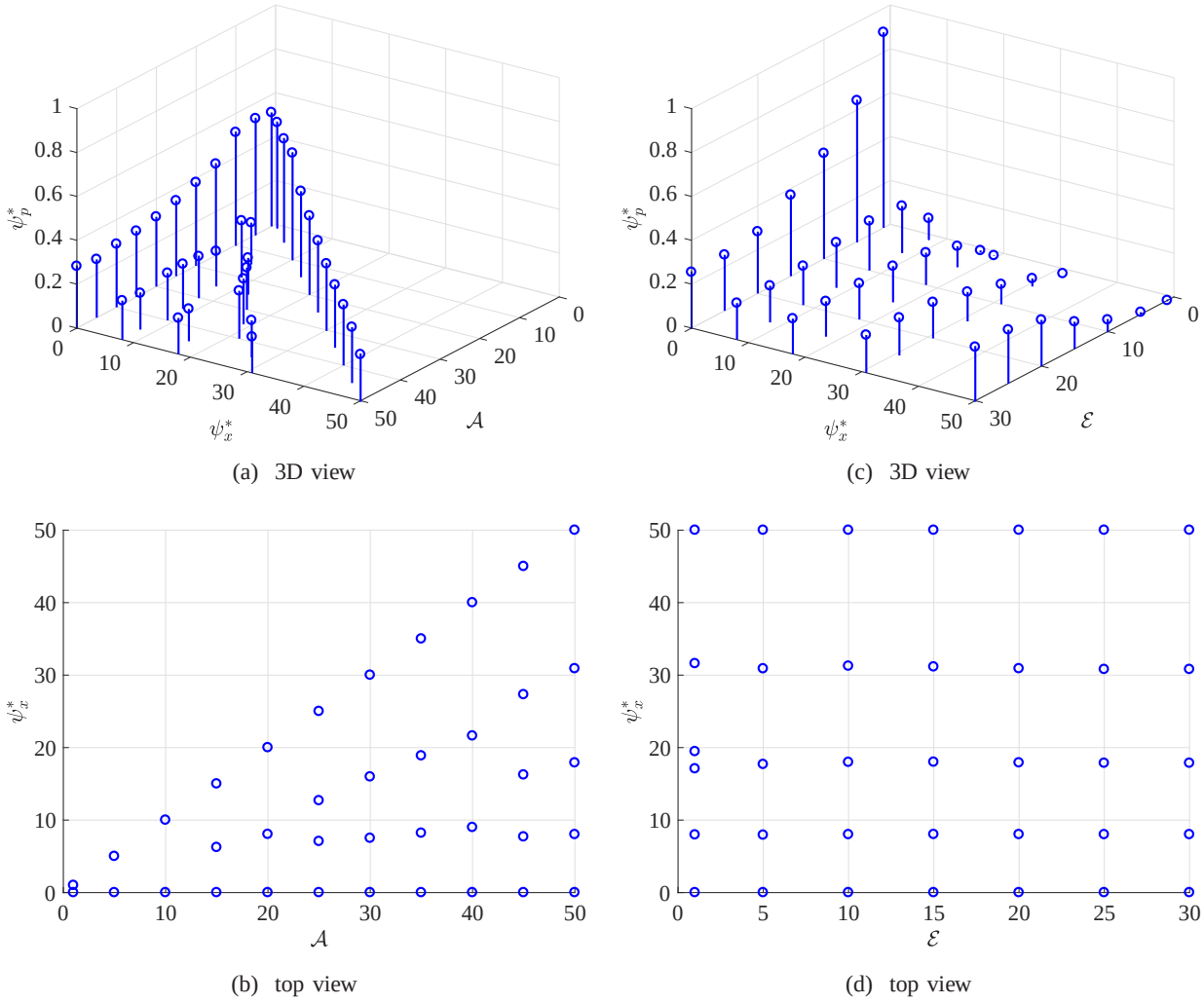
(a) $\delta = 0.1$ (c) $\delta = 0.1$

(b) $\delta = 1$ (d) $\delta = 1$

Fig. 4. Bounds on the information rate of SPAD array for $N_{\mathrm{array}} = 1024$: (a),(b) $\mathcal{E} = 20$; (c),(d) $\mathcal{A} = 50$.

channel capacity remains constant. This is because with the given constraint on the peak power ($\mathcal{A} = 50$), the average power of the optimum input distribution cannot follow the average power constraint closely, i.e., it cannot achieve average powers larger than $0.5\mathcal{A}$.

In addition, for smaller values of $\mathcal{E}$, more mass points are required to achieve the capacity, although the closer the mass points to $\mathcal{A}$, the smaller its corresponding probability, as seen for example in Figs. 5c and 6c for $\mathcal{E} = 1$. Another important observation is that for $N_{\mathrm{array}} = 1024$, if $\delta \leq 1$, the capacity-achieving measure (ψ_x^*, ψ_p^*) does not depend on δ (see Fig. 7). This means that in this large array the impact of the dead time is effectively cancelled. As a numerical example, assuming a typical dead time of 10 ns, using a discrete-time signalling scheme with 5 levels, this array can achieve an information rate of 1.8 bits/channel use with $\mathcal{A} = 50$ and $\mathcal{E} = 20$. This is equivalent to a maximum data rate of 180 Mbits/s with an arbitrarily small BER.



(a) 3D view (c) 3D view

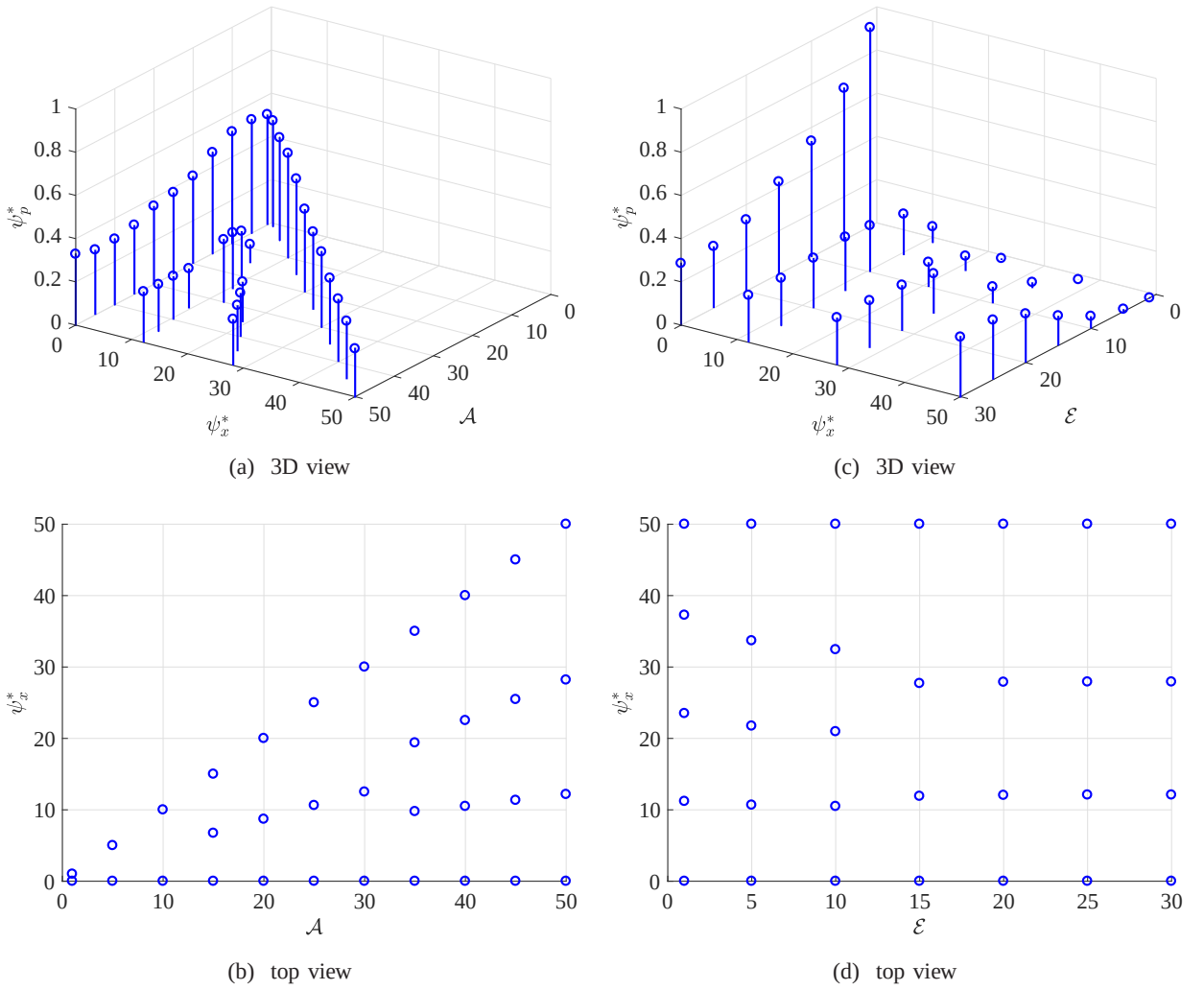
(b) top view (d) top view

Fig. 5. SPAD array capacity-achieving distributions for $N_{\text{array}} = 64$, $\delta = 0.1$ and $K_{\text{b}} = 5$: (a),(b) $\mathcal{E} = 20$; (c),(d) $\mathcal{A} = 50$.

V. SPAD-Based Optical Communication System

In this section, the bit error performance of a SPAD-based optical system with OOK modulation is studied. The OOK modulation is often considered as a benchmark modulation scheme for assessing the error performance of photon counting channels. It is also a special case of the general discrete-time signalling scheme (as studied in the previous section) which has only two intensity levels. Here, the performance of the system is affected by the SPAD dead time and the background counts. In this system, both the primary and the secondary counting losses increase the error probability and limit the maximum achievable data rate of the system.

Consider OOK signalling with a bit time of T seconds, hence, a data rate of $R = 1/T$ bits/s. In OOK, the information bits are transmitted through the intensity of light, where a bit "1" shows the presence of an optical signal pulse and a bit "0" indicates the absence of the signal pulse



(a) 3D view (c) 3D view

(b) top view (d) top view

Fig. 6. SPAD array capacity-achieving distributions for $N_{\text{array}} = 1024$, $\delta = 1$ and $K_{\text{b}} = 10$: (a),(b) $\mathcal{E} = 20$; (c),(d) $\mathcal{A} = 50$.

during each bit interval. Assume that the signal and noise intensities have a uniform spatial distribution over the array area. Denote the optical signal and the background noise photon rates by λ_{s} and λ_{b}, respectively, and define $K_{\text{s}} = \lambda_{\text{s}} T$ and $K_{\text{b}} = \lambda_{\text{b}} T$ as the average signal and background noise counts per bit interval. Let $p_Y^0(y)$ and $p_Y^1(y)$ denote the probabilities of y photocounts, when "0" or "1" are transmitted, respectively.

A. Hypothesis Testing

Consider a classical binary detection process where the hypothesis "H_0" represents the case when a "0" is sent and "H_1" represents the hypothesis that a "1" is transmitted. The receiver attempts to determine the correct bit based upon a single observation of the aggregate number of array photocounts over each bit interval. The detection strategy which minimizes the probability

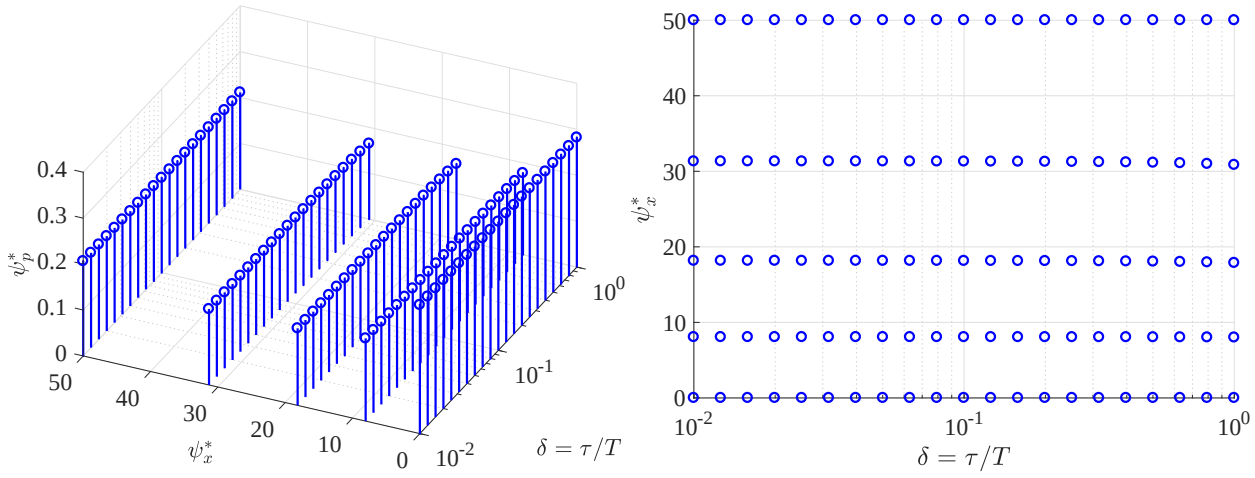


Fig. 7. SPAD array capacity-achieving distributions for $N_{\text{array}} = 1024$, $\mathcal{A} = 50$, $\mathcal{E} = 20$ and $K_{\text{b}} = 5$.

of error for the case of equally likely bits is maximum-likelihood detection. Accordingly, the decision is made based on a likelihood ratio test, defined as:

$$L(y) = \frac{p_Y^1(y)}{p_Y^0(y)} \underset{H_0}{\overset{H_1}{\gtrless}} 1, \tag{34}$$

where $p_Y^1(y) \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $p_Y^0(y) \sim \mathcal{N}(\mu_0, \sigma_0^2)$. The above likelihood ratio test simplifies to a single threshold test for the SPAD array. The optimum threshold value is given by [36]:

$$y_{\text{th}} = \frac{\dfrac{\mu_0}{\sigma_0^2} - \dfrac{\mu_1}{\sigma_1^2} + \sqrt{\left(\dfrac{\mu_0}{\sigma_0^2} - \dfrac{\mu_1}{\sigma_1^2}\right)^2 - \left(\dfrac{1}{\sigma_0^2} - \dfrac{1}{\sigma_1^2}\right)\left[\left(\dfrac{\mu_0^2}{\sigma_0^2} - \dfrac{\mu_1^2}{\sigma_1^2}\right) + 2\ln\left(\dfrac{\sigma_0}{\sigma_1}\right)\right]}}{\dfrac{1}{\sigma_0^2} - \dfrac{1}{\sigma_1^2}}. \tag{35}$$

This threshold can be further approximated as [36]:

$$y_{\text{th}} = \frac{\mu_1\sigma_0 + \mu_0\sigma_1}{\sigma_0 + \sigma_1}. \tag{36}$$

Note that this threshold depends on μ_0, μ_1, σ_0, and σ_1 which are all functions of the average photon rates from the optical signal and background noise, λ_{s} and λ_{b}. Note that λ_{s} and λ_{b} must be known exactly to optimally set the threshold, and this is a technical challenge with the OOK modulation.

B. Bit Error Performance

In OOK demodulation, the number of array photocounts is compared with the threshold y_{th} provided in (35). An error will occur if $y \leq y_{\text{th}}$ when a "1" bit is sent, or if $y > y_{\text{th}}$, when a

"0" bit is sent. The probability of error for equally likely bits is [36], [37]:

$$P_{\mathrm{e}} = \frac{1}{2} \sum_{y=\lfloor y_{\mathrm{th}} \rfloor + 1}^{\infty} p_Y^0(y) + \frac{1}{2} \sum_{y=0}^{\lfloor y_{\mathrm{th}} \rfloor} p_Y^1(y). \tag{37}$$

With the threshold y_{th}, the minimum error probability is ensured. By replacing the discrete variable y with a continuous variable y', P_{e} can be approximated as:

$$\begin{aligned} P_{\mathrm{e}} &\cong \frac{1}{2} \int_{y_{\mathrm{th}}}^{\infty} p_0(y')\, dy' \quad + \frac{1}{2} \int_{0}^{y_{\mathrm{th}}} p_1(y')\, dy' \\ &= \frac{1}{2}\, \mathrm{Q}\left(\frac{y_{\mathrm{th}} - \mu_0}{\sigma_0} \right) + \frac{1}{2}\, \mathrm{Q}\left(\frac{\mu_1 - y_{\mathrm{th}}}{\sigma_1} \right), \end{aligned} \tag{38}$$

where $\mathrm{Q}(x) = 1/\sqrt{2\pi} \int_x^{\infty} \exp(-\alpha^2/2)\, d\alpha$ is the Q-function. In the above equation, $p_0(y')$ and $p_1(y')$ are the continuous approximations of $p_Y^0(y)$ and $p_Y^1(y)$, respectively. With the approximate threshold given in (36), P_{e} simplifies to:

$$P_{\mathrm{e}} \cong \mathrm{Q}\left(\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0} \right). \tag{39}$$

The approximate error probability in (39) suggests that P_{e} depends only on the difference of the signal and background noise mean values. Therefore, any contribution to both μ_1 and μ_0, such as from dark currents, would not effect the $\mu_1 - \mu_0$ term, these will however contribute to the variances. The effective signal-to-noise ratio (SNR) can then be defined as:

$$\gamma = \frac{(\mu_1 - \mu_0)^2}{(\sigma_1 + \sigma_0)^2}. \tag{40}$$

C. Numerical Results and Discussions

In the following we present the bit error performance results of the OWC system with SPAD arrays. Throughout the calculations and simulations, it is assumed that the array elements are identical. In all the figures, the BER results are reported as a function of K_{s} or K_{b} or δ for different array sizes. Also, a data rate of 10 Mbits/s and $C_{\mathrm{FF}} = 1$ are considered. Therefore, dead times of 100 ns and 10 ns correspond to $\delta = 1$ and $\delta = 0.1$, respectively.

In all the numerical and simulation results of this section, the ISI effect is considered, unless stated otherwise. Since closed-form expressions of the ISI-modified photocount distributions are not available, in 'semi-analytical' results, the required PMFs are obtained through Monte Carlo methods and then substituted in the analytical expressions of the error probability. To obtain the required PMFs, when generating a Poisson arrival process and recording the number of

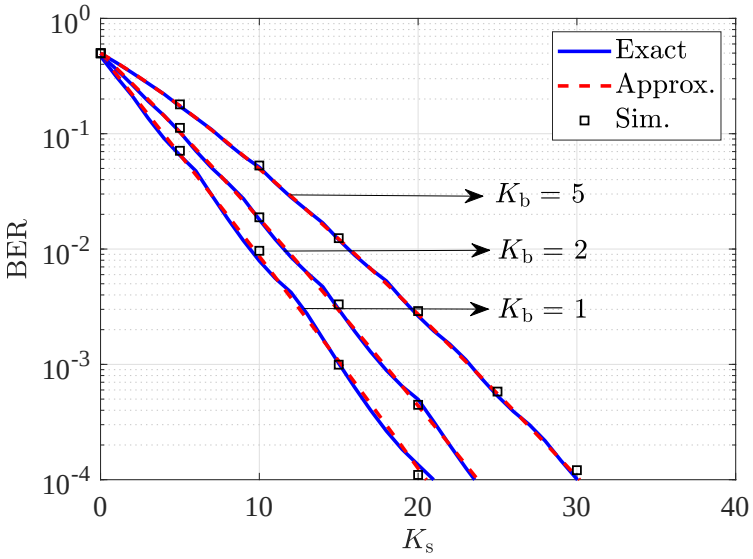


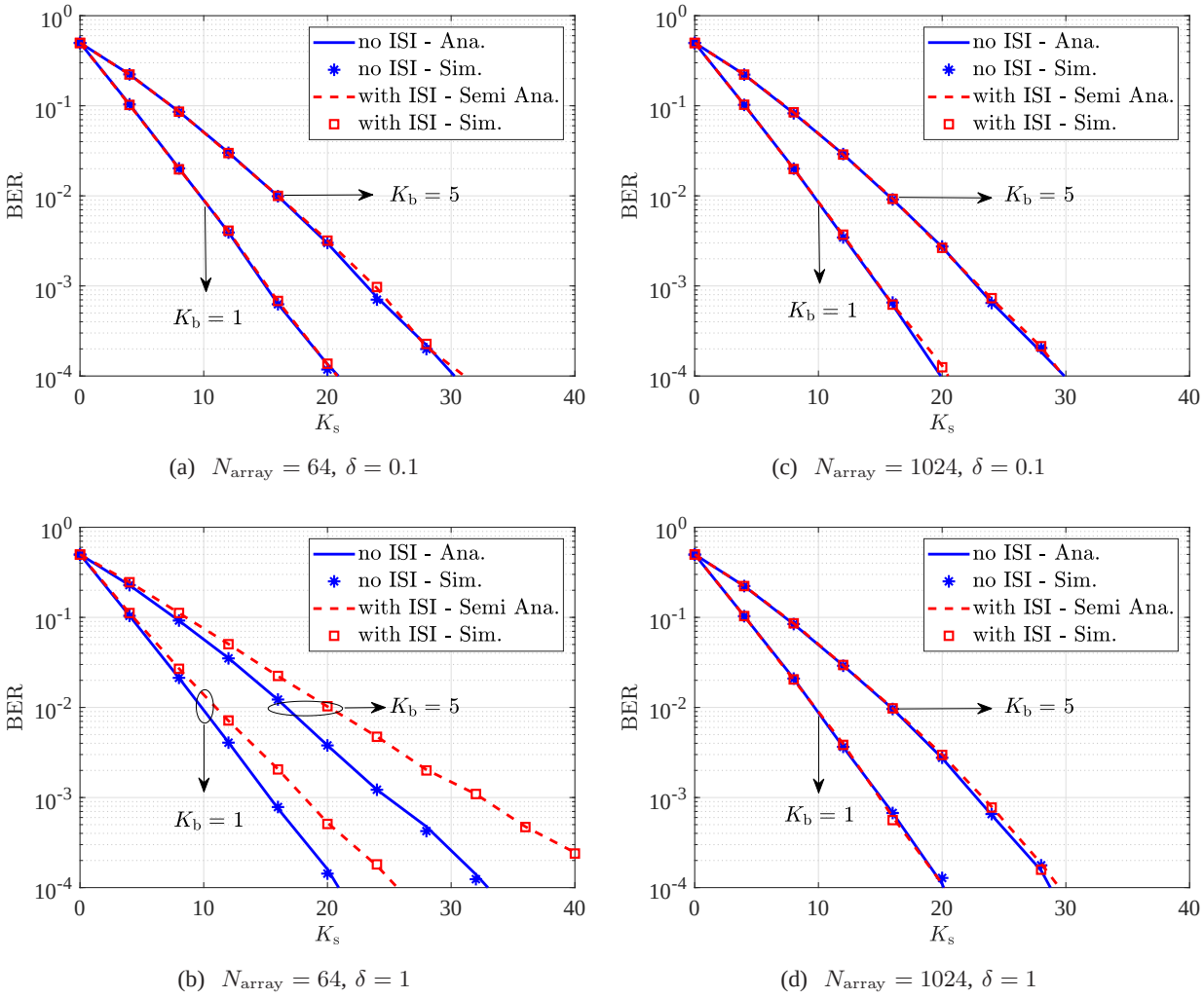
Fig. 8. OOK BER performance with a SPAD array of $N_{\text{array}} = 64$ and $\delta = 0.1$.

photocounts according to the dead time effects, the arrival time of the first photon in each bit interval is compared with the arrival time of the last photocount in the previous bit interval to see whether the first photon is lost or not.

In Fig. 8, the exact error probability given in (37) is numerically evaluated and compared with the approximate error probability of (39) and also the Monte Carlo simulation results for an array of 64 SPADs with dead time ratio of $\delta = 0.1$. Although, the discrete threshold values cause some ripples in the curves, (39) can well approximate the error probability. Hereinafter, the approximate error probability expression given in (39) is adopted for the BER calculations. The BER curves are based on the analytical models, unless otherwise stated.

C.1. ISI

Fig. 9 presents the effect of ISI on the BER of two arrays with 64 and 1024 SPADs. For each array two cases are considered, a short dead time ($\delta = 0.1$) and a long dead time ($\delta = 1$). For $\delta = 0.1$, the ISI is negligible in both arrays. However, the long dead time of $\delta = 1$ results in significant ISI impairment for the array of size $N_{\text{array}} = 64$, while the array of 1024 SPADs is very robust to the ISI. These results indicate that large arrays can suppress the ISI effectively. The reason is that their effective dead time is shorter. Nonetheless, note that the ISI is stronger in higher photon rates. If the photon rate is very high, such that all the SPADs in a large array are saturated, then even the large array may not be able to alleviate the adverse impact of ISI. This case does not take place for the photon rate regimes considered in this paper.



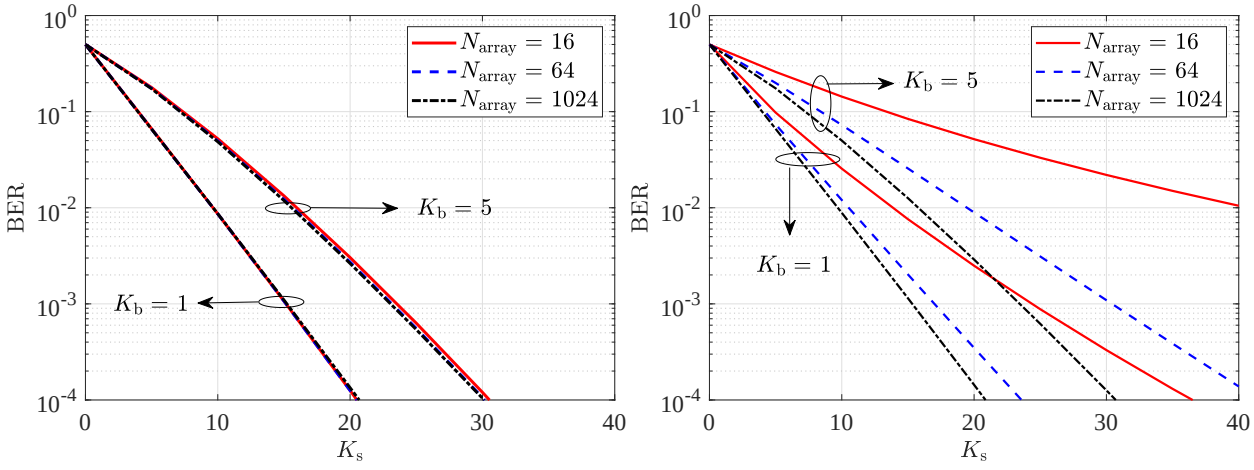
(a) $N_{\text{array}} = 64$, $\delta = 0.1$ (c) $N_{\text{array}} = 1024$, $\delta = 0.1$

(b) $N_{\text{array}} = 64$, $\delta = 1$ (d) $N_{\text{array}} = 1024$, $\delta = 1$

Fig. 9. The effect of ISI on the BER results.

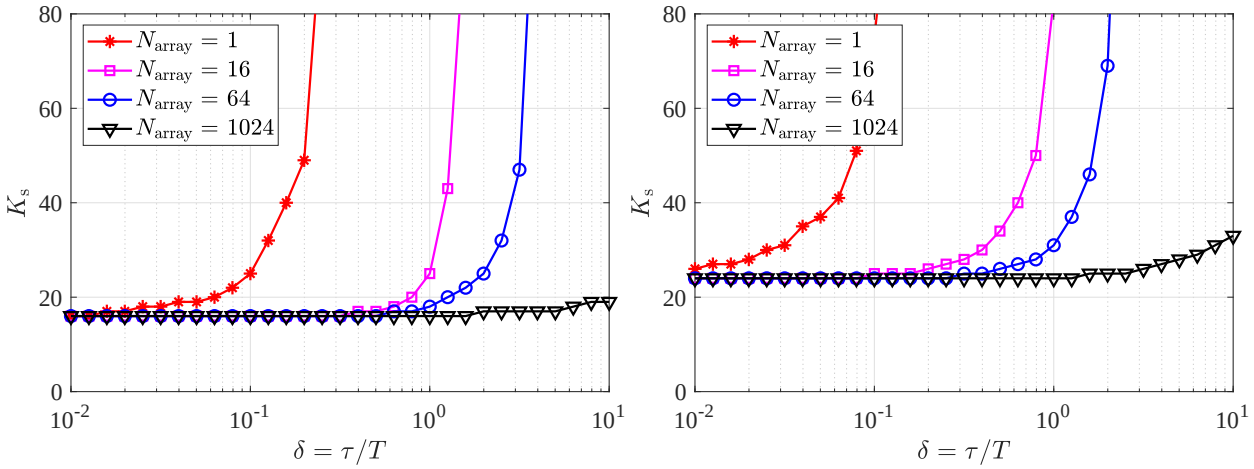
C.2. Array Size

In Fig. 10, the BER of three SPAD arrays with $N_{\text{array}} = 16$, 64, and 1024 are compared for $\delta = 0.1$ and $\delta = 1$. To make a fair comparison, it is assumed that the total sensitive area of the arrays is equal and the average number of signal counts or background noise counts per bit interval is the same for all three arrays. Please note that with this assumption, each SPAD in the array of 1024 SPADs receives lower signal and background noise levels compared to each SPAD in the array of 64 elements. According to Fig. 10a, all arrays perform the same for $K_b = 1$, while for $K_b = 5$ arrays with 64 and 1024 elements slightly outperform the array of 16 SPADs. For $\delta = 1$ as in Fig. 10b, arrays with larger sizes show better error performance for various average background noise levels.



(a) $\delta = 0.1$ (b) $\delta = 1$

Fig. 10. BER results of SPAD array receivers with (a) $\delta = 0.1$ and (b) $\delta = 1$.



(a) $K_b = 1$ (b) $K_b = 5$

Fig. 11. The BER $= 10^{-3}$ contours as a function of the dead time ratio and the average number of photons per bit time required by the SPAD array.

C.3. Dead Time

Fig. 11 shows the effect of dead time on the average number of photons per bit time (i.e., K_s) required by the SPAD array to achieve a particular BER for OOK modulation. In this figure the BER $= 10^{-3}$ contours are displayed as a function of δ and K_s for a single SPAD and three different array sizes. The average background noise count level is $K_b = 1$ in Fig. 11a and $K_b = 5$ in Fig. 11b.

The counting losses due to the dead time are mitigated to some extent when several SPADs

operate in parallel inside an array. It is very unlikely that all the SPADs of an array become inactive at the same time. In low photon rate regimes, if the dead time is short (compared to the bit interval), the use of large size SPAD arrays does not offer considerable performance improvements. In high photon rate regimes, or in the case of long dead times, larger array sizes are required. In such cases, smaller arrays may not achieve the target BER even in higher SNRs, as increasing the optical power leads to the saturation of the SPAD array.

According to Fig. 11a, for shorter dead time durations ($\delta < 0.5$), all three arrays provide BER $\leq 10^{-3}$ with $K_{\rm s} \approx 16$. However, for longer dead times ($\delta > 0.5$), larger arrays require fewer number of photons, i.e., less optical power, to achieve BER $= 10^{-3}$. For instance, for $N_{\rm array} = 16$, if $\delta < 1$, the target BER can be achieved with $K_{\rm s} \approx 16$. As the dead time increases, many of the arriving photons get lost and $K_{\rm s}$ increases very rapidly, such that $K_{\rm s} > 80$ is required for $\delta > 1$. This sharp increase in $K_{\rm s}$ is due to the saturation of the SPAD array. With $N_{\rm array} = 1024$ the limiting effect of dead time is almost eliminated, such that regardless of the value of δ, BER $= 10^{-3}$ is guaranteed with $K_{\rm s} < 20$. Note that depending on the background noise level, the target BER may not be achieved at all; e.g. as in Fig. 11a, for $N_{\rm array} = 64$ and $\delta > 4$.

Consider a target BER of 10^{-3} and denote by $\eta_{\rm array}$ the ratio of required $K_{\rm s}$ for a single SPAD to the required $K_{\rm s}$ for a SPAD array. This ratio can be interpreted as the power gain of the SPAD array compared to the single SPAD for achieving the target BER of 10^{-3}:

$$\eta_{\rm array}\Big|_{\rm BER=10^{-3}} = \frac{K_{\rm s}^{\rm single}}{K_{\rm s}^{\rm array}}\,. \tag{41}$$

Note that the total active area of the array is assumed to be the same as that of the single SPAD. Thus, it is fair to say that the SPAD array requires less power to achieve a target BER. Fig. 12 depicts $\eta_{\rm array}$ for the array of 1024 SPADs. To obtain some of the curves in this figure, minor data extrapolation has been applied.

In Fig. 13, the required $K_{\rm s}$ for achieving some target BERs is plotted as a function of $K_{\rm b}$ for $N_{\rm array} = 64$ and 1024. According to this figure, the minimum required $K_{\rm s}$ to achieve the target BER is constant for $K_{\rm b} \leq 10^{-2}$. However, for $K_{\rm b} > 10^{-2}$, the required $K_{\rm s}$ grows as $K_{\rm b}$ increases. For $N_{\rm array} = 64$, as shown in Fig. 13a, a higher $K_{\rm s}$ is needed to maintain the system performance in the presence of dead time. However, for $N_{\rm array} = 1024$ (see Fig. 13b) the effect of dead time is negligible even for larger values of $K_{\rm b}$.

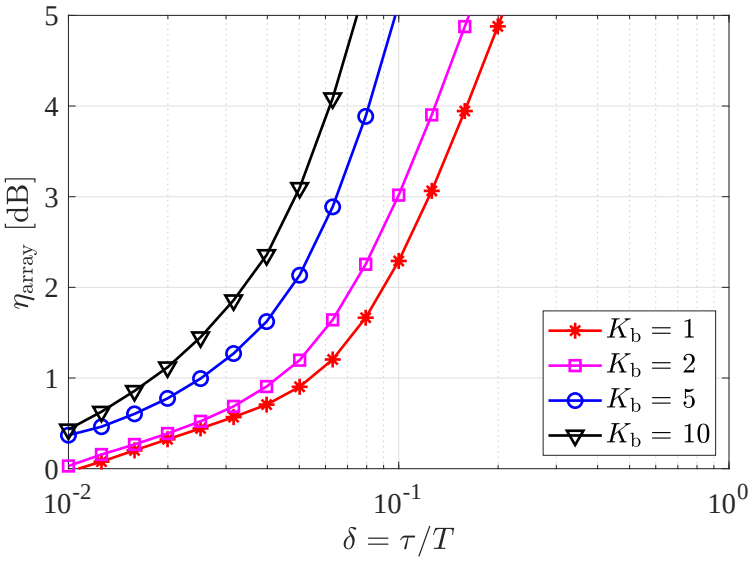
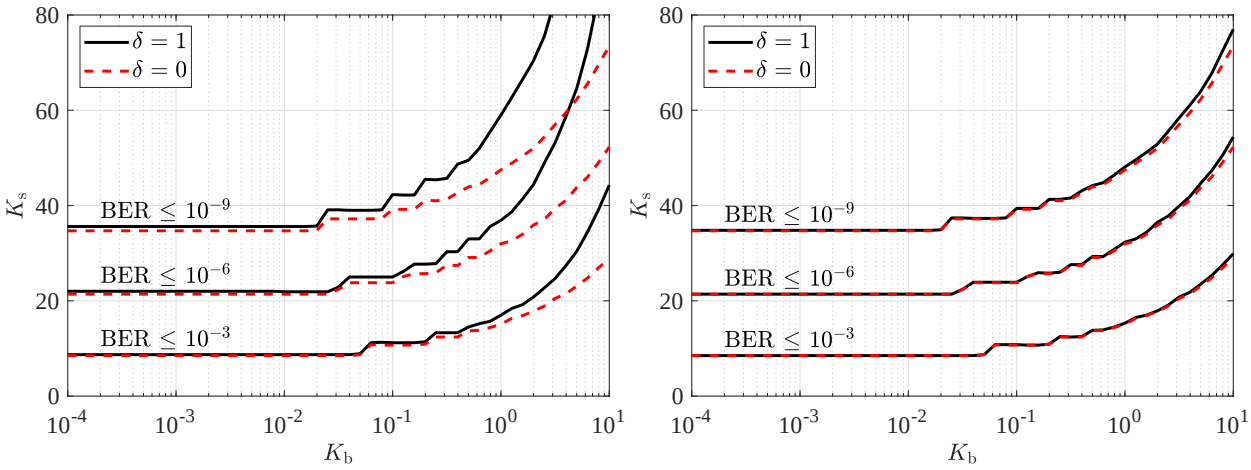


Fig. 12. Power gain of a SPAD array of $N_{\text{array}} = 1024$ for BER $= 10^{-3}$.



(a) $N_{\text{array}} = 64$ (b) $N_{\text{array}} = 1024$

Fig. 13. The BER contours as a function of the average number of signal counts per bit time and average background counts per bit time.

C.4. Background Noise

VI. Conclusion

In this study, the photon counting characteristics, the information rate and the bit error performance of SPAD arrays were studied. It was shown that for sufficiently large SPAD arrays, the photocount distribution can be well approximated by a Gaussian distribution with dead-time-modified mean and variance. The SPAD array was modelled as a communication

channel with a finite memory arising from the dead-time-induced ISI distortions. The SPAD array information rates for a discrete-time signalling scheme were analysed. Using a numerical algorithm, the capacities of two auxiliary channels, subject to average and peak power constraints were evaluated, and the bounds on the information rate of the SPAD array were established. The numerical results demonstrated that in larger arrays the ISI is insignificant. Thus, the SPAD array can be well approximated as a memoryless channel. As such, it is found that for a large array, the optimum input distribution is discrete, consists of a finite set of mass points, and always contains a mass point at zero. In addition, the bit error performance of a SPAD-based OWC system was assessed. The exact and approximate expressions for the error probability of OOK modulation were obtained. The performance results showed that, compared with a single SPAD of the same sensitive area, the SPAD array can tolerate longer dead times maintaining the system performance and requires less signal power to achieve the same probability of error.

The results of this study confirm that by using an array of SPADs, the counting losses arising from the dead time can be mitigated. Hence, the bit error performance and the data rate of the OWC system can be improved.

This study has provided new insights into the application of SPAD receivers for OWCs. It highlights the trade-off between the SPAD photon counting performance and the data rates of OWC systems; in high data rates the existence of dead time causes significant counting losses and thus, significant data loss. The results of this study can be used as a benchmark for evaluating the efficiency of practical SPAD-based optical systems. They are particularly required for designing efficient modulation schemes, and optimizing the device structure and operating conditions to maximize the achievable data rate.

Appendix A
Main Properties of the SPAD FSC Model

Some of the main properties of the SPAD FSC model are presented in this appendix. **According to the law of total probability, the following equation holds for the conditional probability at time n:**

$$\Pr\{y_n|x_n, S_n\} = \sum_{i=1}^{N_S} p^i(y_n|x_n)\Pr\{S_n = i\} \tag{42}$$

where $p^i(y_n|x_n) = \Pr\{y_n|x_n, S_n = i\}$. Since the channels in $\mathcal{S}$ are memoryless:

$$\Pr\{y_n|x_n, S_n, x^{n-1}, S^{n-1}, y^{n-1}\} = \Pr\{y_n|x_n, S_n\} \tag{43}$$

If the x_n's are independent, then:

$$\Pr\{y_n, x_n|S_n, S^{n-1}, x^{n-1}, y^{n-1}\} = \Pr\{y_n, x_n|S_n\} \tag{44a}$$

$$\Pr\{y^n, x^n|S^n\} = \prod_{i=1}^{n} \Pr\{y_i, x_i|S_i\} \tag{44b}$$

$$\Pr\{y_n|S^n, S^{n-1}, y^{n-1}\} = \Pr\{y_n|S_n\} \tag{44c}$$

Appendix B
Proof of Theorem 2

The information rate $I(X^n; Y^n)$ can be written as [39]:

$$I(X^n; Y^n) = H(Y^n) - H(Y^n|X^n), \tag{45}$$

where

$$H(Y^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1}), \tag{46a}$$

$$H(Y^n|X^n) = \sum_{i=1}^{n} H(Y_i|X_i, X^{i-1}, Y^{i-1}). \tag{46b}$$

The summation terms in each of the above equations are nonnegative and monotonically decreasing in i. Thus [40]:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i|Y^{i-1}) = \lim_{n\to\infty} H(Y_n|Y^{n-1}), \tag{47a}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i|X_i, X^{i-1}, Y^{i-1}) = \lim_{n\to\infty} H(Y_n|X_n, X^{n-1}, Y^{n-1}). \tag{47b}$$

The term $H(Y_n|Y^{n-1})$ can be written in terms of $\boldsymbol{\alpha_n}$ as follows:

$$\begin{aligned} H(Y_n|Y^{n-1}) &= \mathbb{E}\left[-\log \Pr\{y_n|y^{n-1}\}\right] \\ &= \mathbb{E}\left[-\log \sum_{i=1}^{N_S} \Pr\{y_n, S_n = i|y^{n-1}\}\right] \\ &= \mathbb{E}\left[-\log \sum_{i=1}^{N_S} \Pr\{y_n|S_n = i, y^{n-1}\}\Pr\{S_n = i|y^{n-1}\}\right] \\ &= \mathbb{E}\left[-\log \sum_{i=1}^{N_S} \Pr\{y_n|S_n = i\}\Pr\{S_n = i|y^{n-1}\}\right] \\ &= \mathbb{E}\left[-\log \sum_{i=1}^{N_S} p^i(y_n)\alpha_n(i)\right]. \end{aligned} \tag{48}$$

Similarly, $H(Y_n|X_n, X^{n-1}, Y^{n-1})$ can be written in terms of $\boldsymbol{\beta_n}$:

$$\begin{aligned}
H(Y_n|X_n, X^{n-1}, Y^{n-1}) &= \mathbb{E}\left[-\log \Pr\{y_n|x_n, x^{n-1}, y^{n-1}\}\right] \\
&= \mathbb{E}\left[-\log \sum_{i=1}^{N_S} \Pr\{y_n, S_n = i|x_n, x^{n-1}, y^{n-1}\}\right] \\
&= \mathbb{E}\left[-\log \sum_{i=1}^{N_S} \Pr\{y_n|S_n = i, x_n, x^{n-1}, y^{n-1}\}\Pr\{S_n = i|x_n, x^{n-1}, y^{n-1}\}\right] \\
&= \mathbb{E}\left[-\log \sum_{i=1}^{N_S} \Pr\{y_n|S_n = i, x_n\}\Pr\{S_n = i|x^{n-1}, y^{n-1}\}\right] \\
&= \mathbb{E}\left[-\log \sum_{i=1}^{N_S} p^i(y_n|x_n)\beta_n(i)\right]. \qquad (49)
\end{aligned}$$

Therefore, according to (45)-(49), the mutual information of the SPAD FSC is given by:

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} I(X^n; Y^n) = &\lim_{n\to\infty} \mathbb{E}\left[-\log \sum_{i=1}^{N_S} p^i(y_n)\alpha_n(i)\right] \\
&- \lim_{n\to\infty} \mathbb{E}\left[-\log \sum_{i=1}^{N_S} p^i(y_n|x_n)\beta_n(i)\right]. \qquad (50)
\end{aligned}$$

Now consider any iid input distribution P_X with the average mutual information of I^* as given in (50). According to Theorem 1, $\boldsymbol{\alpha}_n$ and $\boldsymbol{\beta}_n$ converge to steady distributions $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$, respectively, and y_n be denoted as y^* in the steady state. Thus:

$$I^* = \mathbb{E}\left[-\log \sum_{i=1}^{N_S} p^i(y^*)\alpha^*(i)\right] - \mathbb{E}\left[-\log \sum_{i=1}^{N_S} p^i(y^*|x)\beta^*(i)\right]. \qquad (51)$$

Thus:

$$I_{\text{iid}} \triangleq \max_{P_X} I^*. \qquad (52)$$

If the X_i's are independent, the following inequality holds [50]:

$$\lim_{n\to\infty} \frac{1}{n} I(X^n; Y^n) \geq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_i) = \lim_{n\to\infty} I(X_n; Y_n) = I(X; Y^*). \qquad (53)$$

Recall from Section III-A that the steady state distributions are adopted for the DMC_2. Hence, according to the inequality in (53), for iid inputs:

$$I_{\text{iid}} \geq C_{\text{DMC}_2}. \qquad (54)$$

Now define $\hat{Y}^n = [\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n]$ where $\hat{Y}_i$'s are the ISI-less outputs. At any time instant i, the following equation holds:

$$\Pr\{x_i, \hat{y}_i, y_i\} = p(x_i) \Pr\{\hat{y}_i|x_i\} \Pr\{y_i|\hat{y}_i, x_i\}, \tag{55}$$

where,

$$\Pr\{y_i|\hat{y}_i, x_i\} = \int_0^\tau \Pr\{y_i|\hat{y}_i, x_i, \rho\} f_\rho(\rho) d\rho = \int_0^\tau \Pr\{y_i|\hat{y}_i, \rho\} f_\rho(\rho) d\rho = \Pr\{y_i|\hat{y}_i\}. \tag{56}$$

The second equality holds because y_i is independent of x_i given $\hat{y}_i$ and ρ. By substituting (56) into (55), it is concluded that X_i, $\hat{Y}_i$ and Y_i, and hence X^n, $\hat{Y}^n$ and Y^n form a Markov chain $X^n \longrightarrow \hat{Y}^n \longrightarrow Y^n$. According to the data processing inequality [47], $I(X^n; \hat{Y}^n) \geq I(X^n; Y^n)$. Therefore:

$$\lim_{n\to\infty} \frac{1}{n} I(X^n; Y^n) \leq \lim_{n\to\infty} \frac{1}{n} I(X^n; \hat{Y}^n) \overset{(\star)}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(X_i; \hat{Y}_i) = I(X; \hat{Y}),$$

where in $(\star)$ the equality holds due to the memorylessness of $\hat{Y}_i$'s. The above inequality shows that I_{iid} is upper bounded by the information rate of an ISI-less DMC channel, i.e, DMC_1:

$$I_{\text{iid}} \leq C_{\text{DMC}_1}. \tag{57}$$

ACKNOWLEDGEMENT

Professor Harald Haas gratefully acknowledges the support of this research by the Engineering and Physical Sciences Research Council (EPSRC) under Established Career Fellowship grant, EP/R007101/1, and the financial support of his research by the Wolfson Foundation and the Royal Society. Dr. Majid Safari acknowledges the EPSRC grant ARROW, EP/R023123/1. All authors acknowledge the financial support from EPSRC grant TOWS, EP/S016570/1.

REFERENCES

[1] A. Spinelli and A. Lacaita, "Physics and Numerical Simulation of Single Photon Avalanche Diodes," *IEEE Trans. Electron Devices*, vol. 44, no. 11, pp. 1931–1943, Nov. 1997.

[2] A. Eisele *et al.*, "185 MHz Count Rate, 139 dB Dynamic Range Single-Photon Avalanche Diode with Active Quenching Circuit in 130 nm CMOS Technology," in *Proc. Int. Image Sensor Workshop*, Japan, 2011, pp. 278–281.

[3] A. Gallivanoni, I. Rech, and M. Ghioni, "Progress in Quenching Circuits for Single Photon Avalanche Diodes," *IEEE Trans. Nucl. Sci.*, vol. 57, no. 6, pp. 3815–3826, Dec. 2010.

[4] D. Chitnis and S. Collins, "A SPAD-Based Photon Detecting System for Optical Communications," *IEEE/OSA J. Lightw. Technol.*, vol. 32, no. 10, pp. 2028–2034, May 2014.

[5] D. Chitnis, L. Zhang *et al.*, "A 200 Mb/s VLC Demonstration with a SPAD Based Receiver," in *IEEE Summer Topicals Meeting Series*, Jul. 2015, pp. 226–227.

[6] E. Fisher, I. Underwood, and R. Henderson, "A Reconfigurable 14-bit 60GPhoton/s Single-Photon Receiver for Visible Light Communications," in *Proc. European Solid-State Circuits Conf.*, Bordeaux, France, 2012, pp. 85–88.

[7] ——, "A Reconfigurable Single-Photon-Counting Integrating Receiver for Optical Communications," *IEEE J. Solid-State Circuits*, vol. 48, no. 7, pp. 1638–1650, Jul. 2013.

[8] O. Almer, D. Tsonev *et al.*, "A SPAD-Based Visible Light Communications Receiver Employing Higher Order Modulation," in *Proc. IEEE Global Communications Conf.*, San Diego, CA, Dec. 2015, pp. 1–6.

[9] J. Kosman, K. Moore, H. Haas, and R. K. Henderson, "Distortion Losses of High-Speed Single-Photon Avalanche Diode Optical Receivers Approaching Quantum Sensitivity," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 378, no. 2169, 2020.

[10] J. Kosman, O. Almer, A. V. Jalajakumari, S. Videv, H. Haas, and R. K. Henderson, "60 Mb/s, 2 Meters Visible Light Communications in 1 klx Ambient Using an Unlensed CMOS SPAD Receiver," in *Photonics Society Summer Topical Meeting Series*, Newport Beach, CA, 2016, pp. 171–172.

[11] L. Zhang, H. Chun, G. Faulkner, D. O'Brien, and S. Collins, "A Comparison Between the Sensitivities of VLC Receivers Containing an Off-the-Shelf SPAD Array and an APD," in *Proc. IEEE Photonics Conf.*, Orlando, FL, USA, Oct. 2017, pp. 27–28.

[12] L. Zhang, D. Chitnis *et al.*, "A Comparison of APD-and SPAD-Based Receivers for Visible Light Communications," *IEEE/OSA J. Lightw. Technol.*, vol. 36, no. 12, pp. 2435–2442, 2018.

[13] Y. Li, M. Safari *et al.*, "Single Photon Avalanche Diode (SPAD) VLC System and Application to Downhole Monitoring," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, 2014, pp. 2108–2113.

[14] J. Zhang, L.-H. Si-Ma, B.-Q. Wang, J.-K. Zhang, and Y.-Y. Zhang, "Low-Complexity Receivers and Energy-Efficient Constellations for SPAD VLC Systems," *IEEE Photon. Technol. Lett.*, vol. 28, no. 17, pp. 1799–1802, 2016.

[15] L.-H. Si-Ma, J. Zhang, B.-Q. Wang, and Y.-Y. Zhang, "Energy-Efficient Multidimensional Hellinger Modulation for SPAD-Based Optical Wireless Communications," *Opt. Express*, vol. 25, no. 19, pp. 22 178–22 190, Sep. 2017.

[16] ——, "Hellinger-Distance-Optimal Space Constellations," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 765–768, Apr. 2017.

[17] Y.-D. Zang, J. Zhang, and L.-H. Si-Ma, "Anscombe Root DCO-OFDM for SPAD-Based Visible Light Communication," *IEEE Photon. J.*, vol. 10, no. 2, pp. 1–9, Apr. 2018.

[18] L. Zhang, H. Chun, G. Faulkner, D. O'Brien, and S. Collins, "Efficient Pulse Amplitude Modulation for SPAD-Based Receivers," in *Global LiFi Congress*, Paris, France, Feb. 2018, pp. 1–5.

[19] T. Shafique, O. Amin, M. Abdallah *et al.*, "Performance Analysis of Single-Photon Avalanche Diode Underwater VLC System Using ARQ," *IEEE Photon. J.*, vol. 9, no. 5, pp. 1–11, 2017.

[20] M. Elamassie and M. Uysal, "Performance Characterization of Vertical Underwater VLC Links in the Presence of Turbulence," in *Int. Symp. Communication Systems, Networks & Digital Signal Processing*, 2018, pp. 1–6.

[21] C. Wang, H.-Y. Yu, Y.-J. Zhu, T. Wang, and Y.-W. Ji, "Multi-LED Parallel Transmission for Long Distance Underwater VLC System with One SPAD Receiver," *Elsevier Opt. Commun.*, vol. 410, pp. 889–895, 2018.

[22] ——, "Multiple-Symbol Detection for Practical SPAD-Based VLC System with Experimental Proof," in *Proc. IEEE Global Commun. Conf. Workshops*, Singapore, 2017, pp. 1–6.

[23] C. Wang, H. Yu, Y.-J. Zhu, T. Wang, and Y. Ji, "One Symbol Training Receiver for the SPAD-Based UVLC System," *Applied optics*, vol. 57, no. 20, pp. 5852–5858, 2018.

[24] T. Mao, Z. Wang, and Q. Wang, "Receiver Design for SPAD-Based VLC Systems under Poisson–Gaussian Mixed Noise Model," *Opt. Express*, vol. 25, no. 2, pp. 799–809, 2017.

[25] J. Xiang, Y. Jia *et al.*, "Channel Likelihood Correction for Photon-Counting Array Receivers in the Presence of Dead Time and Jitters," *Opt. Express*, vol. 26, no. 3, pp. 2848–2856, 2018.

[26] Y. Li, M. Safari, R. Henderson, and H. Haas, "Optical OFDM with Single-Photon Avalanche Diode," *IEEE Photon. Technol. Lett.*, vol. 27, no. 9, pp. 943–946, May 2015.

[27] ——, "Nonlinear Distortion in SPAD-Based Optical OFDM Systems," in *Proc. IEEE Global Commun. Conf. Workshops*, San Diego, CA, 2015.

[28] T. Hamza, M.-A. Khalighi, S. Bourennane, P. Leon, and J. Opderbecke, "On the Suitability of Employing Silicon Photomultipliers for Underwater Wireless Optical Communication links," in *Int. Symp. on Communication Systems, Networks & Digital Signal Processing*, Jul. 2016, pp. 1–5.

[29] M. A. Khalighi, T. Hamza *et al.*, "Underwater Wireless Optical Communications Using Silicon Photo-Multipliers," *IEEE Photonics J.*, vol. 9, no. 4, pp. 1–10, Aug. 2017.

[30] D. Zou, C. Gong, K. Wang, and Z. Xu, "Characterization of a Practical Photon Counting Receiver in Optical Scattering Communication," in *Proc. IEEE Global Commun. Conf.*, Singapore, 2017.

[31] ——, "Characterization on Practical Photon Counting Receiver in Optical Scattering Communication," *IEEE Transactions on Communications*, 2018.

[32] E. Sarbazi and H. Haas, "Detection Statistics and Error Performance of SPAD-based Optical Receivers," in *Proc. IEEE 26th Ann. Int. Symp. Personal, Indoor, and Mobile Radio Communications*, Hong Kong, China, Sep. 2015, pp. 830–834.

[33] E. Sarbazi, M. Safari, and H. Haas, "Statistical Modeling of Single-Photon Avalanche Diode Receivers for Optical Wireless Communications," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 4043–4058, Apr. 2018.

[34] ——, "The Impact of Long Dead Time on the Photocount Distribution of SPAD Receivers," in *Proc. IEEE Global Communications Conf.*, Abu Dhabi, UAE, Dec. 2018, pp. 1–6.

[35] ——, "On the Information Transfer Rate of SPAD Receivers for Optical Wireless Communications," in *Proc. IEEE Global Communications Conf.*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[36] ——, "Photon Detection Characteristics and Error Performance of SPAD Array Optical Receivers," in *Proc. IEEE 4th Int. Workshop on Optical Wireless Communications*, Istanbul, Turkey, Sep. 2015, pp. 132–136.

[37] R. M. Gagliardi and S. Karp, *Optical Communications*, 2nd ed. New York: Wiley, 1995.

[38] C. C. Chen, "Effect of Detector Dead Time on the Performance of Optical Direct Detection Communication Links," *TDA Progress Rep.*, vol. 42–93, pp. 146–154, Mar. 1988.

[39] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[40] A. J. Goldsmith and P. P. Varaiya, "Capacity, Mutual Information, and Coding for Finite-State Markov Channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[41] D. M. Arnold, "Computing Information Rates of Finite-State Models with Application to Magnetic Recording," Ph.D. dissertation, ETH Zurich, Zurich, Switzerland, 2003.

[42] S. Shamai, L. H. Ozarow, and A. D. Wyner, "Information Rates for a Discrete-Time Gaussian Channel with Intersymbol Interference and Stationary Inputs," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1527–1539, 1991.

[43] S. Ghavami, R. S. Adve, and F. Lahouti, "Information Rates of ASK-Based Molecular Communication in Fluid Media," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 1, no. 3, pp. 277–291, 2015.

[44] J. G. Smith, "The Information Capacity of Amplitude- and Variance-Constrained Scalar Gaussian Channels," *Inf. Ctrl.*, vol. 18, no. 3, pp. 203–219, 1971.

[45] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-Achieving Probability Measure for Conditionally Gaussian Channels with Bounded Inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, Jun. 2005.

[46] S. M. Moser, "Capacity Results of an Optical Intensity Channel with Input-Dependent Gaussian Noise," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 207–223, Jan. 2012.

[47] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

[48] A. V. Fiacco and G. P. McCormick, "The Sequential Unconstrained Minimization Technique (SUMT) without Parameters," *Operations Research*, vol. 15, no. 5, pp. 820–827, 1967.

[49] L. Jia, "Modified Sequential Unconstrained Minimization Technique Methods for Constrained Optimization." Ph.D. dissertation, University of Illinois, Urbana, IL, USA, 1995.

[50] W. Hirt, "Capacity and Information Rates of Discrete-Time Channels with Memory," Ph.D. dissertation, ETH Zurich, Zurich, Switzerland, 1988.

Response to Reviewers' Comments for Manuscript TCOM-TPS-19-1459.R1 "The Bit Error Performance and Information Transfer Rate of SPAD Array Optical Receivers"

Elham Sarbazi, Majid Safari and Harald Haas

The authors would like to express their sincere gratitude to the editor and the anonymous reviewers for their insightful and constructive comments, which have doubtlessly improved the quality of the manuscript. We hope that the paper can now be accepted for publication.

1 Reviewer 2

The answer to my question 2 is not satisfactory. The given statement in the reply and also (42) still seem wrong. To use a very clean notation, the law of total probability says the following:

$$\Pr\{Y=y\}=\sum_i \Pr\{Y=y,S=i\}=\sum_i \Pr\{S=i\}\Pr\{Y=y|S=i\} \quad (1)$$

The left-hand side of (42) now is also conditioned on $X=x$, i.e., everything needs to be conditioned on $X=x$:

$$\begin{aligned}\Pr\{Y=y|X=x\}&=\sum_i \Pr\{Y=y,S=i|X=x\}\\&=\sum_i \Pr\{S=i|X=x\}\Pr\{Y=y|X=x,S=i\}\end{aligned} \quad (2)$$

So, I'm missing the term $\Pr\{S=i|X=x\}$ and I do not have a conditioning on S on the left-hand side. Please explain and fix!

We would like to thank the reviewer for this detailed clarification. We have double checked this expression and agree with the reviewer about the missing term $\Pr\{S_n=i\}$. **We have checked all the steps of the proofs given in appendices and confirm that the rest of the expressions are all correct (e.g., equations (48) and (49) use the same law and are both correct). As a matter of fact, (42) is only presented as one property of the channel and has not been used in the proofs.**

In addition, as requested by the reviewer, in the following we rewrite the conditional probability given in (42). We first drop the conditioning on x_n:

$$\begin{aligned}\Pr\{y_n|S_n\} &= \frac{\Pr\{y_n, S_n\}}{\Pr\{S_n\}} \\ &= \frac{\sum_{i=1}^{N_S} \Pr\{y_n, S_n = i\}}{\Pr\{S_n\}} \\ &= \frac{\sum_{i=1}^{N_S} \Pr\{y_n|S_n = i\}\Pr\{S_n = i\}}{\Pr\{S_n\}} \\ &\overset{(\star)}{=} \sum_{i=1}^{N_S} \Pr\{y_n|S_n = i\}\Pr\{S_n = i\}\end{aligned} \tag{3}$$

Recalling from Sec. III in the revised manuscript, the state space is defined as $\mathcal{S} = \{1, 2, \ldots, N_S\}$ and $S_n \in \mathcal{S}$. The equality in $(\star)$ holds because $\Pr\{S_n\} = \sum_{i=1}^{N_S} \Pr\{S_n = i\} = 1$. Next, we consider conditioning on x_n:

$$\begin{aligned}\Pr\{y_n|x_n, S_n\} &= \frac{\Pr\{y_n, S_n|x_n\}}{\Pr\{S_n|x_n\}} \\ &= \frac{\sum_{i=1}^{N_S} \Pr\{y_n, S_n = i|x_n\}}{\Pr\{S_n|x_n\}} \\ &\overset{(\star\star)}{=} \frac{\sum_{i=1}^{N_S} \Pr\{y_n|x_n, S_n = i\}\Pr\{S_n = i\}}{\Pr\{S_n\}} \\ &= \sum_{i=1}^{N_S} \Pr\{y_n|x_n, S_n = i\}\Pr\{S_n = i\}\end{aligned} \tag{4}$$

The equality in $(\star\star)$ holds because the channel state S_n is independent of x_n. **In response to this comment, (42) in the revised manuscript has been updated.**